

UNIVERSIDAD EXTERNADO DE COLOMBIA  
FACULTAD DE ADMINISTRACIÓN DE EMPRESAS  
MAESTRIA EN GERENCIA ESTRATEGICA DE TECNOLOGIAS DE LA  
INFORMACIÓN (VII Promoción)

ALINEAMIENTO ESTRATÉGICO DE *BIG DATA* EN EL INSTITUTO COLOMBIANO  
DE CRÉDITO EDUCATIVO Y ESTUDIOS TÉCNICOS EN EL EXTERIOR MARIANO  
OSPINA PÉREZ - ICETEX

ANDREA LÓPEZ MATEUS

DIRECTOR DE TESIS: Ing. FRAN ERNESTO ROMERO ÁLVAREZ

BOGOTÁ  
MAYO 2020

APROBACIÓN PARA TRABAJOS DE GRADO

---

Firma de Jurado

---

Firma de Jurado

BOGOTÁ  
MAYO 2020

## Contenido

Glosario .....	8
Introducción .....	10
Capítulo 1. Planteamiento de problema .....	12
1. Antecedentes .....	12
1.1. Problema de investigación.....	14
1.2. Preguntas de investigación .....	15
1.2.1. .... Primaria	15
1.2.2. .... Secundaria	16
1.3. Objetivo general .....	16
1.4. Objetivos específicos.....	16
1.5. Alcance .....	16
1.6. Limitaciones .....	17
1.7. Justificación.....	17
Capítulo 2. Marco contextual.....	19
2.1. Características de la entidad elegida.....	19
2.1.1. Historia .....	20
2.1.2. Productos .....	21
2.1.3. Enfoque estratégico .....	24
Capítulo 3. Marco conceptual .....	25
3.1. Revisión de literatura.....	25
3.2. Arquitectura empresarial basado en <i>Big Data</i> .....	27
3.4 Proceso de desarrollo de Oracle de arquitectura empresarial.....	29
3.5. SAVI-BIGD Marco estratégico .....	36
3.3 Modelos de negocios basados en datos .....	38
3.6 Guía Técnica MinTIC – Gobierno de Datos .....	41
3.6.1 Descripción .....	41
3.6.2 Beneficios .....	41
3.6.3 Flujo de pasos del gobierno de datos.....	42
3.7. Metodologías de analítica basados en <i>Big Data</i> – Metodología CRISP-DM .....	43

3.7.1 Fase I. Comprensión del negocio.....	44
3.7.2 Fase II. Comprensión de datos.....	44
3.7.3 Fase III. Análisis de datos y selección de características .....	44
3.7.4 Fase IV. Modelado.....	45
3.7.5 Fase V. Evaluación (obtención de resultados).....	45
3.7.6 Fase VI. Despliegue (puesta en producción) .....	46
3.8. <i>Big Data</i> : casos de estudio aplicables en el contexto .....	46
3.8.1 Consejo Nacional de Política Económica y Social -CONPES 3920 .....	46
3.8.2 Aplicación de <i>Big Data</i> en la educación, minería de datos y aprendizaje de análisis. .....	47
Capítulo 4. Metodología de la investigación .....	49
4.1 Tipo de investigación .....	49
4.2. Técnica de recolección de datos .....	51
4.3. Instrumentos y Fuentes.....	52
4.4. Instrumentos .....	52
4.4.1 Encuesta.....	52
4.4.2 Cuestionario .....	53
4.5 Identificación de población .....	53
4.5.1 Muestra bases de datos (BD) ICETEX.....	53
4.6. Descripción de la muestra .....	58
4.7. Diseño de los instrumentos.....	58
4.8. Aplicación de instrumentos .....	59
4.9. Tabulación .....	59
Capítulo 5. Análisis de instrumento y propuesta .....	60
5.1. Aplicación de la encuesta .....	60
5.2. Propuesta .....	61
5.3. Estrategia para propuesta marco alineado en <i>Big Data</i> .....	61
5.4. Fase 1: Visión estratégica.....	61
5.4.1. Aplicación y resultados de la encuesta de la siguiente manera .....	62
5.5. Fase 2: Diseño de hoja de ruta para <i>Big Data</i> .....	66
5.6. Fase 3: Generación de Objetivos Estratégicos de <i>Big Data</i> .....	67
5.7. Fase 4: Identificación de fuentes de datos .....	68
5.7.1. Fuentes de búsqueda y posible integración.....	70

5.8. Fase 5: Implementación de <i>Big Data</i> .....	71
5.9 Plan Financiero.....	72
Conclusiones y recomendaciones .....	76
Lista de referencias .....	77
Anexo 1. Preguntas – encuesta .....	81
Anexo 2. Cronograma.....	84

## Índice de tablas

Tabla 1. Reporte últimos 5 años aplazamientos semestres 1 y 2 de estudiantes.....	12
Tabla 2. Solicitudes registradas para créditos 2019-1 y 2019-2 .....	13
Tabla 3. Discriminación por estado de solicitudes de líneas de crédito para pregrado 2019-1 y 2019-2 .....	13
Tabla 4. Líneas de crédito educativo pregrado .....	22
Tabla 5. Líneas posgrado ICETEX .....	23
Tabla 6. Características determinantes para implementar Big Data .....	28
Tabla 7. Fase 1 Contexto y alcance del negocio .....	30
Tabla 8. Fase 2. Establecer una versión de arquitectura .....	30
Tabla 9. Fase 3 Evaluar estado actual .....	32
Tabla 10. Fase 4. Establecer estado futuro.....	33
Tabla 11. Fase 5. Desarrollar hoja de ruta .....	34
Tabla 12. Fase 6. Establecer gobernanza sobre la arquitectura.....	35
Tabla 13. Tabla deserción por semestre 2016.....	56
Tabla 14. Porcentaje de deserción por departamento y ciclo propedéutico .....	56
Tabla 15. Deserción anual por área y nivel de formación.....	58
Tabla 16. Tabulación resultados encuesta.....	59
Tabla 17. Arquitecturas basadas en datos .....	60
Tabla 18. Integrantes dirección de tecnología.....	60
Tabla 19. Equipo de trabajo .....	67
Tabla 20. Análisis económico del sector.....	72
Tabla 21. Porcentaje recursos .....	73
Tabla 22. Proyección a cinco años.....	74

## Índice de figuras

Figura 1. Líneas de crédito recursos propios 2019-1 .....	14
Figura 2. Organigrama directivo ICETEX.....	19
Figura 3. Organización dirección de tecnología .....	20
Figura 4. Valores Institucionales ICETEX .....	21
Figura 5. Porcentaje de datos de regiones desarrolladas.....	26
Figura 6. Fusión estrategia empresarial y TI.....	36
Figura 7. Estrategia fases SAVI-BIGD marco estratégico .....	37
Figura 8. Importancia del dato .....	42
Figura 9. Pasos gobierno de datos.....	43
Figura 10. Ciclo de metodología CRISP-DM.....	44
Figura 11. Ciclo de metodología.....	45
Figura 12. Enfoque de proyectos .....	50
Figura 13. Métodos mixtos .....	51
Figura 14. Estructura crédito ICETEX.....	54
Figura 15. Evolución créditos nuevos.....	55
Figura 16. Usuarios activos por estrato y género .....	55
Figura 17. Fase visión estratégica .....	62
Figura 18. Fase implementación mapa de ruta.....	66
Figura 19. Fase generación objetivos estratégicos .....	67
Figura 20. Objetivos estratégicos de tecnología .....	68
Figura 21. Fase identificación fuentes de datos .....	69
Figura 22. Pasos para desarrollar e implementar la gestión del ciclo de vida del dato.....	69
Figura 23. Fuentes de datos.....	70
Figura 24. Manejo actual de los datos.....	71

## Glosario

**Analítica:** Análisis, reflexión, observación de grandes datos permite mejorar la segmentación de mercados para orientar ofertas y productos e innovar en los modelos de negocios y de producción, así como crear nuevos productos (combinando la producción en masa con la personalización) y modelos de servicios empresariales y gubernamentales. Además de aumentar la transparencia y la eficiencia, posibilitan un mejor y oportuno análisis del desempeño de todo tipo de variables y ajustar estructuras y comportamientos en tiempo real (CEPAL, 2017, 27).

**Big Data:** término que describe el gran volumen de datos, tanto estructurados como no estructurados, que inundan un negocio día a día. Pero no es la cantidad de datos lo que es importante, sino lo que hacen las organizaciones con los datos. Los datos grandes se pueden analizar en busca de ideas que conducen a mejores decisiones y movimientos comerciales estratégicos

**Ciclo propedéutico:** la Ley 749 de 2002 introduce en el sistema educativo la formación por ciclos con carácter propedéutico, específicamente en las áreas de ingenierías, la tecnología de la información y la administración. Estas son unidades interdependientes, complementarias y secuenciales; para continuar en el proceso de formación a lo largo de la vida, en este caso particular, en el pregrado. En consecuencia, un ciclo propedéutico se puede definir como una fase de la educación que le permite al estudiante desarrollarse en su formación profesional siguiendo sus intereses y capacidades.

**CRISP-DM:** *Cross- Industry Standard Practice for Data Mining*. Es una metodología de enfoque multinacional, basado en estándares para describir documentar y mejorar continuamente los procesos de extracción de datos asociado almacenamiento de datos e inteligencia de negocios (Ville, 2001).

Proceso estándar entre la industria de la minería de datos (CRISP-DM), modelo de minería de datos no propietaria, documentado y de libre acceso creado por IBM y Daimler AG, que consta de seis fases para la elaboración e implementación de su proceso.

**Dato:** La palabra dato se deriva del latín DTUM, y su significado es “lo que se da”. Durante el tiempo se ha visto la necesidad de clasificar los datos y algunos especialistas los han trabajado para hacer cálculos estadísticos, matemáticos entre otros, que han facilitado la toma de decisiones en diferentes corporaciones, de igual manera se ha utilizado en diferentes áreas de conocimiento para hacer mejoras de productos e innovaciones.

**DevOps:** Metodología de desarrollo de software que se centra en la comunicación, colaboración e integración entre desarrollos de software y los profesionales de sistemas en las tecnologías de la información, aplicando buenas prácticas tecnológicas.

**Git Hub:** Plataforma de desarrollo colaborativo de software para alojar proyectos utilizando el sistema de control de versiones, este código se almacena de forma pública.

**ICETEX:** Instituto colombiano de crédito educativo y estudios técnicos en el exterior  
Mariano Ospina Pérez – ICETEX

**ICFES:** Instituto colombiano para el fomento de la educación superior

**Inteligencia de negocios:** Es una herramienta que le permite a las organizaciones tomar decisiones basadas en información precisa y oportuna del estudio de los datos, garantizando la creación del conocimiento necesario para escoger una alternativa conveniente para el éxito de una empresa.

**IPC:** Índice del Precio al Consumidor

**Jenkins:** Es una multiplataforma que actualmente se ha consolidado como una de las herramientas más usadas en la actualidad para realizar las tareas de integración continua.

**Minería de datos:** Es el proceso de analizar los datos desde diferentes perspectivas y resumiéndolos en información útil, en información que se puede utilizar para aumentar los ingresos, reducir los costos, o ambas cosas.

Es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

**MinTIC:** Ministerio de Tecnologías de la Información y las comunicaciones

**SPADIES:** Sistema para la Prevención y Análisis de la Deserción en las Instituciones de Educación Superior

**TI:** Tecnologías de Información

**TyT:** Técnicas y Tecnológicas

**TRM:** Tasa Representativa del mercado

## Introducción

Los datos son un elemento valioso para las organizaciones porque representan el insumo principal para la elaboración de estrategias de mercadotecnia que permitan tomar decisiones y generar valor a las empresas. Herramientas como Data Mining, Big Data, Inteligencia de Negocios e Inteligencia Artificial actualmente determinan los comportamientos o las tendencias en un contexto o mercado específico.

Si bien es cierto que existen restricciones enmarcadas en la normatividad colombiana (Ley 1581 de 2012) frente al derecho fundamental que tienen los ciudadanos para la protección de sus datos personales, también es posible aprovechar la información, con previa autorización del titular, para una finalidad de uso estadístico.

El mercado del siglo XXI está diseñado para que los datos suministrados por los usuarios en diferentes plataformas de Internet –como las redes sociales– se puedan aprovechar para generar patrones que permitan obtener información de preferencias de algún bien o servicio. De acuerdo con la clasificación de estos datos es posible orientar al usuario para tomar una decisión y realizar determinada inversión, por ejemplo, la publicidad digital; claro está que la orientación que las plataformas envían al usuario mediante noticias, sugerencias, correos, avisos, entre otros, responde a un análisis computacional previo de los datos. Gracias a esta clasificación de datos se obtienen diferentes opciones para estudiar, comprar, viajar y otras oportunidades que apuntan a una adquisición de bienes y servicios personalizada.

Las búsquedas que realizan los usuarios en Internet, relacionadas con el ICETEX, generan información que queda reportada en bases de datos (número de visitas, preguntas frecuentes, número de usuarios, entre otros), lo que es motivo suficiente para que la Oficina Comercial y la Oficina de Mercadeo de esta entidad de créditos estudiantiles aproveche los datos para crear campañas dirigidas a medios de comunicación, y así solucionar dudas habituales, evitar el desplazamiento de los clientes y promover otros servicios adicionales que consoliden una cultura empresarial que cubra todos los frentes de acción.

Existen compañías como Datacredito –una central de información financiera y crediticia de personas y compañías– que tienen convenio con las entidades bancarias de Colombia para realizar consultas de los estados financieros de los usuarios; este es un ejemplo representativo donde se evidencia el uso acertado de datos debido a que no es necesario dirigirse físicamente a la entidad. Datacredito entrega información financiera de los usuarios para, posteriormente, realizar un estudio de riesgo y según los resultados proporcionar un crédito. La entidad tiene calificaciones cuantitativas, ya sean positivas o negativas, que les permite a los prestamistas entregar créditos de distintas modalidades; por ejemplo: compra de vehículos, inmuebles, tecnología, telecomunicaciones y afiliaciones, entre otras.

Más empresas en Colombia están utilizando Machine Learning, Inteligencia Artificial, Big Data o Data Mining con el fin de mejorar la competitividad y la innovación en sus organizaciones, ya que esta información es la Fuente: principal y la clave para realizar negocios y tomar decisiones. Con este contexto, valiéndose de los datos digitales autorizados por los usuarios, se quiere efectuar una investigación para diseñar un modelo de asignación

de recursos en ICETEX para determinar qué ciudadanos realmente necesitan de un crédito. En esa medida, se pretende identificar las regiones de Colombia que requieren de más profesionales para contribuir al desarrollo de las ciudades y así guiar a la compañía crediticia para tener una idea precisa de cuál puede ser la mejor opción de un estudiante al momento de acceder a un programa académico de educación superior (Portafolio, 2017)

## Capítulo 1. Planteamiento de problema

### 1. Antecedentes

El ICETEX es una entidad del Estado que contribuye a la educación superior del país por medio de la asignación de recursos propios, administración de fondos y alianzas estratégicas con diferentes entidades, aliados y constituyentes, con el propósito de guiar a los jóvenes colombianos que quieren estudiar y no tienen el dinero para acceder a la educación superior, ni la capacidad de endeudamiento que exigen las entidades financieras.

El ICETEX administra recursos que asigna la Nación para estudios en educación superior, no solamente a jóvenes mayores de edad, sino también a los que se gradúan entre los 15 y 17 años del colegio. El objetivo es que no pierdan tiempo y puedan estudiar inmediatamente por medio de un crédito educativo, situación que no ocurre con otras entidades financieras, mientras que en el ICETEX es una característica especial. Esta población representa una gran cantidad de datos que pueden utilizarse para producir reportes dirigidos a diferentes áreas de la entidad. Por ejemplo, los jóvenes que inician sus estudios por primera vez presentan un porcentaje de deserción que se ve reflejado cada semestre que abren las convocatorias para renovaciones y nuevos créditos. En la tabla 1 se evidencian las cifras de los últimos cinco años con respecto a estudiantes que han abandonado sus estudios de educación superior.

*Tabla 1. Reporte últimos 5 años aplazamientos semestres 1 y 2 de estudiantes*

<b>Estudiantes que abandonan sus estudios y quedan con el crédito y la deuda</b>					
<b>Estudiantes que abandonan en I y II Semestre</b>	<b>2015 I-II</b>	<b>2016 I-II</b>	<b>2017 I-II</b>	<b>2018 I-II</b>	<b>2019 I-II</b>
APLAZADO	9%	8%	9%	23%	17%
DESERTO	10%	10%	8%	6%	3%
Total, general	19%	18%	17%	29%	20%

Fuente: Elaboración propia de descarga reportes ICETEX 2019

Ahora bien, al revisar estos cinco años se puede determinar que aproximadamente el 21% de la deserción corresponde a recursos asignados con características no conocidas. Para el 2019 de las 171.415 solicitudes registradas, el 20% corresponde a deserción, aplazamiento, o desistieron a la solicitud, en la tabla 2 se realiza la discriminación de los usuarios registrados que se inscribieron a una línea de crédito.

*Tabla 2. Solicitudes registradas para créditos 2019-1 y 2019-2*

<b>Estado de solicitudes pregrado y posgrado</b>		
1. Aprobados	35.521	26.376
2. En Proceso	38.268	33.804
3. No Toma Crédito	25.228	12.218
<b>Total, general</b>	<b>99.017</b>	<b>72.398</b>

Fuente: Elaboración propia de descarga reportes ICETEX 2019

Los datos se publican cada seis meses. Para el primer semestre de cada año, enero – junio, la demanda o solicitudes son más altas que para el segundo semestre, julio - diciembre, como se puede evidenciar en la tabla 3.

*Tabla 3. Discriminación por estado de solicitudes de líneas de crédito para pregrado 2019-1 y 2019-2*

<b>Estado</b>	<b>2019-1</b>	<b>2019-2</b>
Cancelado por usuario	20.061	284
En proceso	38.268	33.804
Créditos aprobados	35.521	26.373
Rechazados	3.630	11.824
Pendientes en revisión	1.537	113
<b>Total, general</b>	<b>99.017</b>	<b>72.398</b>

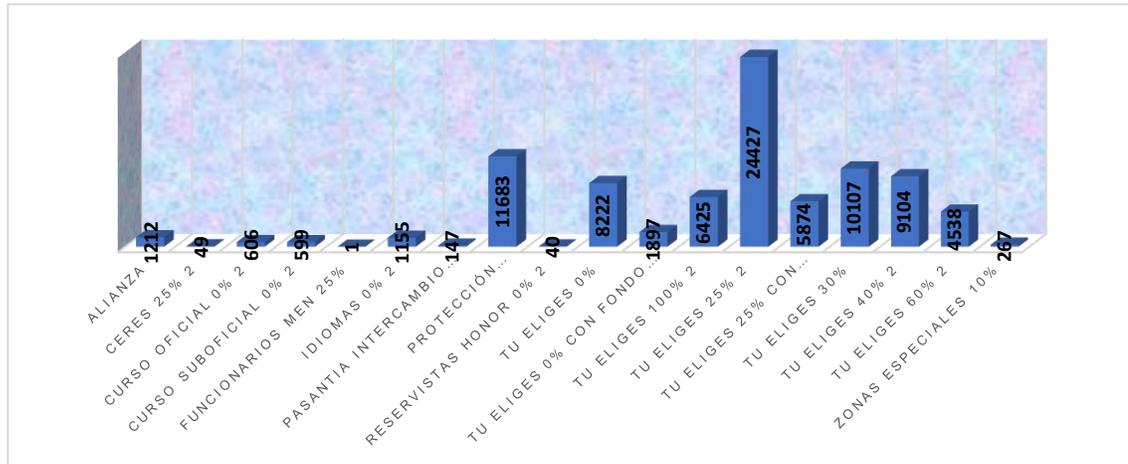
Fuente: Elaboración propia de descarga del reporteador ICETEX 2019

Los valores de tablas anteriores corresponden a las líneas de crédito de recursos propios, los cuales generan el 50% de los datos almacenados en el ICETEX. El otro 50% son de créditos condonables de Fondos en Administración; es decir, convenios con otras entidades y constituyentes que asignan recursos para promover la educación superior.

Las líneas de crédito que hacen parte de los recursos propios del ICETEX se denominan ‘Crédito Tú Eliges’. Estas tienen características y condiciones que se adaptan a las diferentes posibilidades del estudiante y su contexto económico. Así mismo, existen líneas especiales para los usuarios ubicados en regiones apartadas de las capitales y para las personas que pertenecen a población vulnerable.

Las líneas de crédito se cobran una vez que el estudiante culmine su programa académico o renuncie a finalizar sus estudios. En la figura 1 se observan las solicitudes de estudiantes para la ‘Convocatoria 2019-1: líneas de crédito para pregrado’.

Figura 1. Líneas de crédito recursos propios 2019-1



Fuente: Elaboración propia descarga reportes ICETEX-2019

Como se mencionó anteriormente, el ICETEX también cuenta con otras líneas para cursar estudios de educación superior como las líneas de Fondos en Administración; es decir, convenios con otras entidades del estado y entidades privadas, que son subsidiadas y, por lo general, son créditos condonables. Estos créditos no hacen parte de las cifras de estudiantes que desertan o abandonan, debido a que son líneas subsidiadas y tienen otras características que generan reportes distintos.

### 1.1. Problema de investigación

El ICETEX cuenta actualmente con diferentes Fuentes de información para luego ser analizadas en los diferentes reportes que requiere la entidad y tienen una carga operativa ya que la mayoría se maneja por medio reportes en Excel o archivos ETL, es decir, archivos planos que se cargan en los aplicativos bajo una funcionalidad programada. A partir de estos se realiza un análisis para obtener la información de una característica en especial, como pueden ser: cantidad de inscritos en una línea de crédito, porcentaje de inscritos en líneas de fondos, estratos socioeconómicos, programas con más personas inscritas, estudiantes que desertaron o abandonaron su crédito, entre otras.

Si se toman los porcentajes de la característica de la deserción, según las cifras del Sistema para la Prevención de la Deserción de la Educación Superior (SPADIES), se reporta que un 51,72% de los estudiantes que no tienen crédito abandonan sus estudios de educación superior. Por su parte, para los usuarios que acceden a un crédito con el ICETEX, la cifra es solamente del 21%, lo que indica que la deserción es más baja en estudiantes que pueden acceder a un crédito para financiar su educación, siendo este un reporte que se genera cada semestre.

Como puede observarse, es alto el porcentaje de deserción estudiantil, una cifra que tiene datos de características mencionadas anteriormente y en este momento no se pueden discriminar para plantear un seguimiento o una estrategia. Ahora bien, la falta de profesionales se ve reflejada en las dificultades laborales y competitivas de los sectores económicos de la sociedad, y en una menor cantidad de personal capacitado que requieren las regiones del país para aumentar sus niveles de desarrollo. Por si fuera poco, no se tiene en cuenta en las cifras a las personas que no culminan su programa de estudio superior, pero que sí permanecen con la deuda de un crédito educativo, aumentando los porcentajes de estudiantes con cartera.

Por otro lado, el número de estudiantes que se gradúan de universidades con poco reconocimiento, no logran posicionarse laboralmente y contribuyen a aumentar la tasa de desempleo. Así mismo, es fundamental identificar a los estudiantes que culminaron su educación con la ayuda de un crédito educativo, sin embargo, todavía no consiguen ubicarse laboralmente, razón por la que tienen que pagar el crédito bajo dificultades económicas por falta del ingreso fijo de un empleo.

Características como el desempleo de estudiantes graduados con crédito ICETEX, son de los análisis que se pueden revisar para entender el comportamiento de estos datos y sus familias. Buscando estrategias que permitan orientar a futuro a nuevos estudiantes que son los que contribuyen a aumentar el número de profesionales en las diferentes áreas de conocimiento; tarea que todavía no logra identificarse a fondo por no contar con herramientas o estrategias que analicen una gran cantidad de estos datos.

Entre las problemáticas que afronta el ICETEX y afecta a los usuarios están la deserción académica, la cartera o las cuentas por cobrar a estudiantes graduados y no graduados, la empleabilidad de los usuarios de crédito, su ubicación geográfica, las necesidades de un país con pocos profesionales, entre otras. Son dificultades que manifiestan una alerta y requieren de una estrategia para la clasificación y organización de datos, de manera que se conviertan en activo imprescindible.

Si bien es cierto que existen datos de usuarios externos que son beneficiarios de un crédito educativo y pueden generar información valiosa para el progreso de sus familias y del país, también están los usuarios internos que utilizan estos datos para realizar los diferentes reportes, informes de metas y planes que los entes de control y la institución necesitan. Estos informes se realizan bajo consultas en las bases de datos o, si son pequeños, por medio de hojas electrónicas. Una tarea que todavía no se logra identificar por no contar con herramientas o estrategias que realicen estos análisis.

## **1.2. Preguntas de investigación**

### **1.2.1. Primaria**

¿Cómo alinear estratégicamente *Big Data* en el ICETEX a fin de aprovechar los datos para el diseño de políticas y estrategias en beneficio de sus usuarios y la entidad?

### 1.2.2. Secundaria

¿Cómo la adopción de *Big Data* le permitirá al ICETEX transformar en conocimiento útil sus datos endógenos y exógenos?

¿Cuáles son las falencias y dificultades a superar para la adopción de *Big Data* en el ICETEX?

¿Cómo soportar las necesidades tecnológicas en cuanto a la gobernanza de los datos del ICETEX dentro del plan estratégico para las nuevas tecnologías de la información?

## 1.3. Objetivo general

Proponer una metodología de *Alineamiento Estratégico de Big Data*, cuya puesta en práctica le permita al ICETEX consolidar una plataforma tecnológica adecuada para sustentar sus procesos estratégicos y operativos.

## 1.4. Objetivos específicos

- Identificar la metodología de alineamiento estratégico de *Big Data* que mejor se ajuste a las características y necesidades del ICETEX
- Diseñar una propuesta de aplicación de la metodología adoptada
- Identificar el conjunto de recomendaciones y buenas prácticas que permitirán una aplicación efectiva y exitosa de *Big Data* al interior del ICETEX.
- Diagnosticar y modelar el entorno tecnológico a través de un ejercicio de alineamiento basado en los marcos de referencia de estrategia de *Big Data* que permita identificar la mejor opción para la estrategia del ICETEX.

## 1.5. Alcance

Para lograr cumplir con este proyecto, se pretende implantar estándares y buenas prácticas, que permitan alinear los datos que reposan en el ICETEX para mejorar y aprovechar la información, definir responsabilidades al personal y generar confianza en la toma de decisiones y en la planeación de la estrategia. La calidad y la integridad de los datos de los usuarios, de las regiones, de los programas académicos y de quienes finalizan sus estudios ayudarán a dicho objetivo.

Las áreas de negocio como la Vicepresidencia de Crédito y Cobranza, Vicepresidencia de Fondos en Administración, la Vicepresidencia de Operaciones y Tecnología, la Oficina de Relaciones Internacionales, la Oficina de Riesgos y la Vicepresidencia Financiera crean reportes e indicadores para el manejo de los datos y la asignación de los recursos del ICETEX. Estas áreas serán beneficiadas, pues el buen manejo de los datos minimiza tareas y procesos.

Este estudio pretende agrupar, medir y analizar los datos históricos de las líneas de crédito del ICETEX, cuyo objetivo es adjudicar créditos educativos a personas de bajos recursos que quieren cursar un programa de educación superior. Estas líneas que tienen la capacidad para autosostenerse, son reembolsables y con cartera activa. Estas corresponden al 64% y el 36% a las líneas de fondos.

## **1.6. Limitaciones**

La falta de calidad de datos que tiene la entidad hace que las actividades propuestas sean demoradas o cambie las fechas del cronograma. Por otro lado, llegar al cambio cultural y el entendimiento desde el área de tecnología no es tan sencillo. Todas las aplicaciones y beneficios que tiene *Big Data*, se deben hacer de una de forma progresiva, pero de igual manera se inicia con analítica de áreas específicas por la falta de expertos en el tema.

Si bien es cierto que la planeación estratégica se realiza cada año, para este año se tendrá en cuenta un alineamiento estratégico desde *Big Data* no se realizara en un 100%, y como se mencionó será progresivo y en áreas específicas, para que sean modelo y buena práctica en las siguientes planeaciones.

Para la fase de identificación de Fuentes se iniciará con la integración de la Registraduría, Icfes, Ministerio del Interior y Unidad de Víctimas, las demás fuentes serán en una segunda fase y que involucren las demás áreas.

## **1.7. Justificación**

Una buena gestión de datos es clave para un buen servicio. Para ello, es necesario controlar los datos desde un enfoque donde participen los implicados como la alta gerencia y el área de tecnología, creando una coordinación que represente y entienda las partes interesadas. Es importante tener en cuenta que los datos son un activo que pueden llegar a ser eficientes en las estrategias tomadas por las entidades.

Una entidad como el ICETEX, con aproximadamente 744.000 beneficiarios activos, debe pensar en una estrategia que permita analizar el comportamiento de los datos para mejorar y aprovechar la información que ahí reposa, unificar los procesos e integrarlos con todas las áreas, sistematizar reportes que permitan entregar información que generen ideas para mejorar los procesos y procedimientos que hagan entregas más eficientes. Además, como entidad financiera de carácter especial, cuya finalidad es promover créditos educativos con recursos propios y de terceros, debe buscar estrategias que permitan llegar a la mayor cantidad de colombianos

Para analizar las necesidades y posibles preferencias de estudio de los colombianos, se puede usar la información histórica del ICETEX para buscar alternativas de financiamiento en educación superior, ya que la educación pública no abastece toda la demanda y se hace necesario cubrir esta demanda a las personas que no lograron ingresar a universidades públicas, pero desean continuar sus estudios por medio de un crédito educativo.

Esta información almacenada sin ningún tratamiento y sin lograr identificar posibles beneficios, familias que acceden a créditos educativos, el fortalecimiento en el ámbito laboral y la reducción de pobreza podrán ser presentados.

Esto se puede lograr con la implementación de iniciativas en el uso de *Big Data*, *BI/Data Warehousing*, para facilitar la migración de datos, cumplir con la normativa, reducir el riesgo empresarial, aumentar la flexibilidad empresarial y la agilidad de negocio, mejorar la accesibilidad de los datos, contar con la integridad de datos y responder a las nuevas tecnologías.

## Capítulo 2. Marco contextual

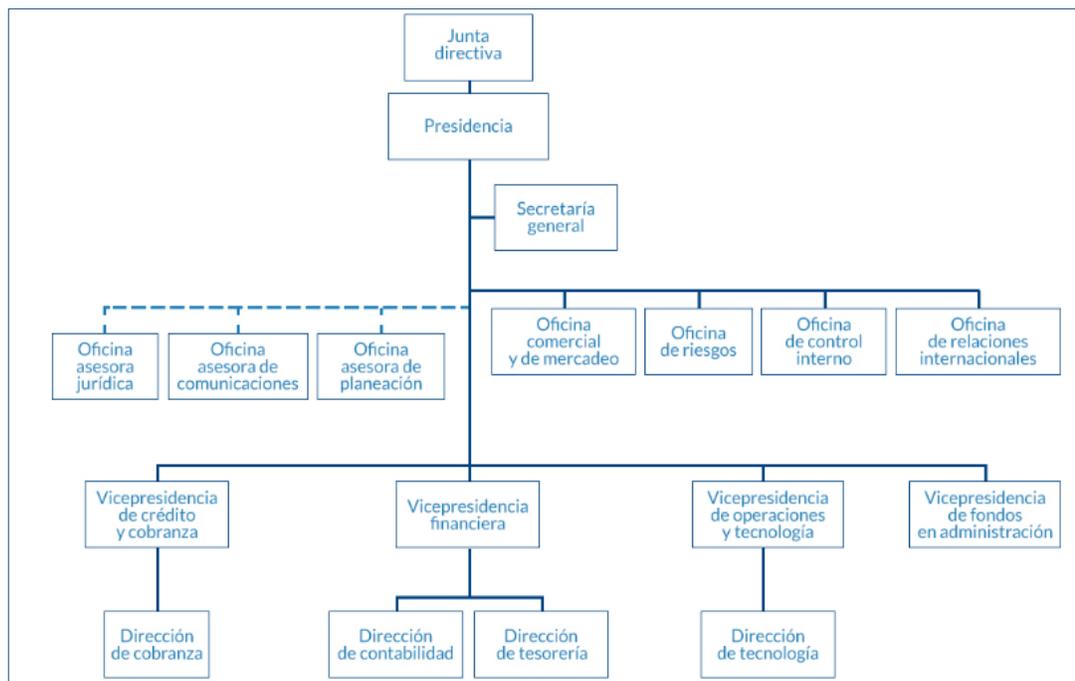
### 2.1. Características de la entidad elegida

Para este capítulo, la Fuente: de información central será la información que reposa en la página del ICETEX acerca de su historia, visión, misión y características. Como se ha mencionado anteriormente, el ICETEX es una entidad del Estado adscrita al Ministerio de Educación Nacional que administra recursos de la Nación, según el presupuesto estipulado por la Presidencia de la República, para apoyar la Educación Superior a través del otorgamiento de créditos y así participar en el crecimiento y desarrollo económico y social del país.

El ICETEX, en convenio con la comunidad internacional, facilita el acceso a becas estudiantiles en el exterior. Adicionalmente, las empresas están adaptando nuevas tecnologías en su proceso educativo, ofreciendo estudios en línea, por lo que es necesario actualizar el modelo de otorgamiento de créditos y pensar en posibilidades para cubrir este nuevo mercado.

En la figura 2, se puede observar cómo está establecida la estructura del ICETEX.

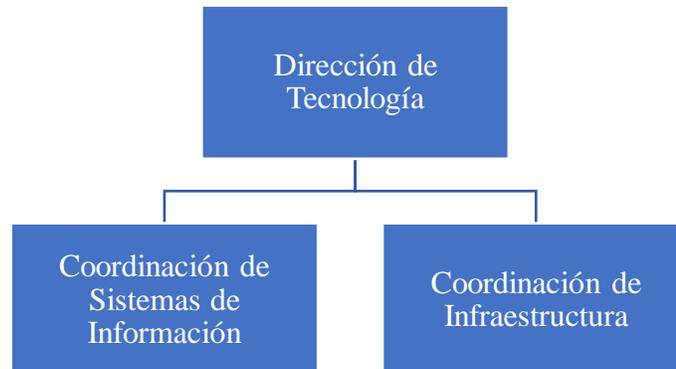
*Figura 2. Organigrama directivo ICETEX*



Fuente: (ICETEX, s.f.b.)

La estructura de la entidad está dividida en vicepresidencias, direcciones y oficinas; estas últimas, a su vez, en coordinaciones. En la figura 3, se discrimina la dirección de tecnología, área que pretende proponer una estrategia para alinear a los objetivos de la entidad basados en *Big Data* o en tecnologías que puedan estudiar y analizar los datos que reposan en las bases de datos.

*Figura 3. Organización dirección de tecnología*



Fuente: Elaboración propia adoptada del ICETEX

### 2.1.1. Historia

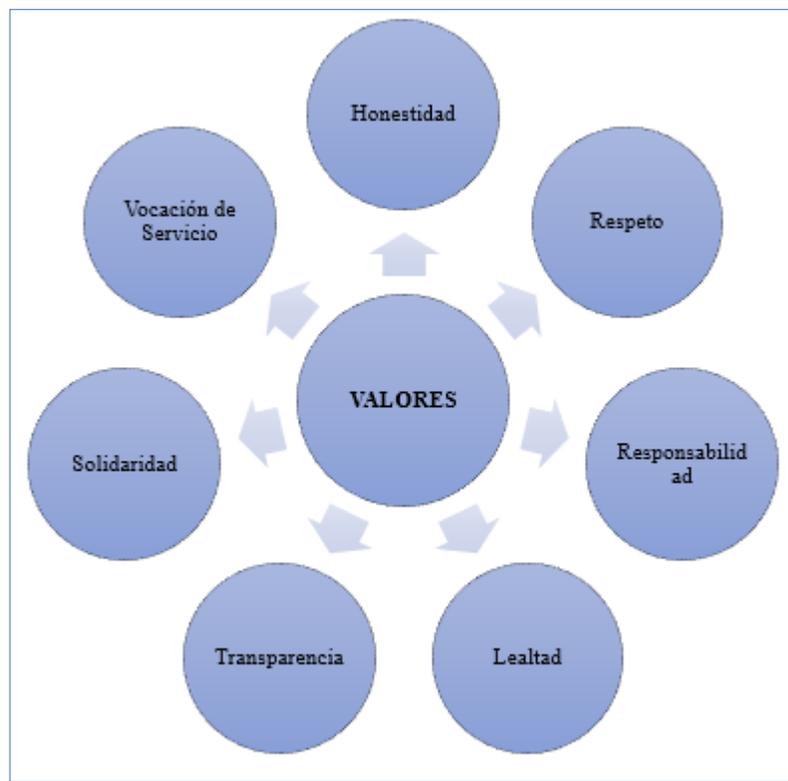
Según la ley 1002 del 30 de diciembre de 2005 del Congreso de la República de Colombia, el ICETEX fue fundado en la ciudad de Bogotá hace 69 años por el doctor Betancur Mejía. Con el Artículo 2 se le da otorga naturaleza especial, con personería jurídica y vinculada al Ministerio de Educación Nacional.

Una de sus características es fomentar socialmente la educación superior, priorizando la población de bajos recursos económicos y a aquellos con mérito académico en todos los estratos sociales, a través de mecanismos financieros que hagan posible el acceso y la permanencia. Las becas son apoyos de carácter nacional e internacional, cumpliendo algunos criterios de cobertura, calidad y pertinencia educativa, en condiciones de equidad territorial. Igualmente, otorga subsidios para apoyar a los estudiantes de estratos 1, 2 y 3.

Uno de los proyectos presidenciales para el 2025 es posicionar al ICETEX como una de las entidades más queridas por los colombianos, pues contribuye significativamente a la transformación social y a una Colombia educada. Se espera que, para aquel momento, la entidad sea una de los cinco líderes internacionales en gestión de recursos para el fomento de la educación, con eficiencia organizacional, orientación al cliente, innovación y apoyo tecnológico. El ICETEX no está sometido a régimen de encajes ni a inversiones forzosas y tampoco podrá ser obligado a destinar recursos de su portafolio para adquirir títulos de deuda pública.

Entre los objetivos estratégicos del ICETEX, está descrito contribuir al acceso de la educación superior en Colombia, llegando a los lugares más apartados del país, optimizando procesos administrativos con la ayuda de las tecnologías de la información y generando estrategias para asignar recursos a los estudiantes que quieren disminuir el déficit educativo y laboral del país, con el propósito de aumentar la competitividad a nivel nacional e internacional. Los colaboradores del ICETEX fortalecen los valores institucionales para el cumplimiento de las metas. En la figura 4, se mencionan los siete valores institucionales que crean pertenencia a cada uno de los funcionarios.

*Figura 4. Valores Institucionales ICETEX*



Fuente: Elaboración propia adoptada de ICETEX

### 2.1.2. Productos

Actualmente el ICETEX otorga créditos, subsidios y beneficios a través de los siguientes productos:

- Líneas de crédito tradicional

Las líneas tradicionales son aquellas líneas de crédito denominadas “Tú Eliges” (pregrado y posgrado) y líneas especiales para pregrado. Estos créditos se dividen en porcentajes que son pagados mientras el beneficiario o usuario estudia su programa académico en pregrado. Estas líneas las administra la Vicepresidencia de Crédito y Cobranza con apoyo de las áreas de tecnología, financiera y la oficina de riesgos. A continuación, en la tabla 4, se describen las características de cada línea.

Tabla 4. Líneas de crédito educativo pregrado

LÍNEA	NOMBRE DE LA LÍNEA	ESTRATO	REGIONES	PUNTAJE ICFES
TU ELIGES	Tú Eliges 0%	1, 2 o 3	14 ciudades < 30,39 Resto Urbano < 30,73 Rural < 22,19	Puntaje Saber 11 $\geq$ 300 o Promedio $\geq$ 3,4
	Tú Eliges 25%	1, 2 o 4	N/A	Puntaje Saber 11 $\geq$ 270 o Promedio $\geq$ 3,4
	Tú Eliges 30%	N/A	N/A	Puntaje Saber 11 $\geq$ 260 o Promedio $\geq$ 3,4
	Tú Eliges 40%	N/A	N/A	Puntaje Saber 11 $\geq$ 240 o Promedio $\geq$ 3,4
	Tú Eliges 60%	N/A	N/A	Puntaje Saber 11 $\geq$ 240 o Promedio $\geq$ 3,4
	Tú Eliges 100%	N/A	N/A	Puntaje Saber 11 $\geq$ 240 o Promedio $\geq$ 3,4
Militares	OFICIALES	N/A	N/A	Puntaje Saber 11 $\geq$ 240 o Promedio $\geq$ 3,4
	SUBOFICIALES	N/A	N/A	Puntaje Saber 11 $\geq$ 240 o Promedio $\geq$ 3,4
	RESERVISTAS DE HONOR	N/A	N/A	Puntaje Saber 11 $\geq$ 240 o Promedio $\geq$ 3,4
Líneas Especiales	Mas colombiano que Nunca	1, 2, 3	14 ciudades < 30,39 Resto Urbano < 30,73 Rural < 22,19	Puntaje Saber 11 $\geq$ 210 o Promedio $\geq$ 3,4
	Protección Constitucional	N/A	N/A	Puntaje Saber 11 $\geq$ 210 Indígenas $\geq$ 200 Promedio $\geq$ 3,4

Fuente: Elaboración propia

Para las líneas de postgrado existen dos opciones que aplican según las condiciones del beneficiario, pueden estudiar un posgrado en el país o en el exterior de igual manera de las posibilidades o elecciones de cada uno. En la tabla 5 se puede observar las características de estas líneas.

Tabla 5. Líneas posgrado ICETEX

LINEA			REQUISITOS MÍNIMOS
TÚ ELIGES	Tú Eliges 20% Posgrado en el país	Se realiza el 20% de pago en época de estudio y el saldo al finalizar.	*Requisitos mínimos: Estar admitido en un programa que cuente con reconocimiento oficial del Ministerio de Educación. *Tener un deudor solidario aceptado por El ICETEX. *Tener historial financiero y crediticio favorable. *Si como solicitante has tenido o tienes un crédito con El ICETEX, debes haber cancelado mínimo el 51% de la deuda. *Autorizar a El ICETEX para consultar, registrar y reportar antecedentes crediticios en las entidades idóneas creadas para tal fin. *Tasa de interés: IPC + 8 puntos porcentuales (Efectivo Anual).
	Tú Eliges 0% Posgrado en el exterior	No realiza pagos en época de estudio, una vez finaliza y se gradúa paga el 100%, en el doble de tiempo que duro el programa académico.	

Fuente: Elaboración propia

Este estudio se enfocará en estudiantes que quieren acceder a estudiar un pregrado, puesto que quienes realizan un posgrado deben tener como requisito inicial ser profesional en algún programa académico. Los estudiantes de posgrado también generan estadísticas que pueden ser útiles para realizar convenios y aportar a las necesidades de las diferentes industrias del país.

- Fondos en administración

Es una línea de negocio del ICETEX encargada de administrar de manera conjunta y exitosa los fondos constitutivos por terceros para invertir en la educación. La Ley 30 de 1992 y la Ley 1002 de 2005 faculta al ICETEX como administrador de recursos. Esta línea es administrada por la Vicepresidencia de Fondos en administración.

- Becas

Es una línea de negocio que realiza otorgamiento de becas internacionales para colombianos y otorgamiento de becas para extranjeros en Colombia, por medio de convenios internacionales y entidades del Estado que aportan para la educación superior. Esta línea la administra la oficina de Relaciones Internacionales.

### 2.1.3. Enfoque estratégico

Actualmente, el ICETEX está trabajando en su modernización tecnológica. Para ello, una de sus tareas es la capacitación a sus funcionarios sobre prácticas tecnológicas que se implementan en diferentes mercados mediante el análisis de datos. Las capacitaciones hacen parte del buen uso de las tecnologías y facilitan el trabajo en cuanto a la interpretación de los datos para optimizar algunos procesos y la calidad de atención a los usuarios.

Entre los planes propuestos por la actual administración, se pretende la implementación de uno de los modelos de analítica, inteligencia de negocios, minería de datos o *Big Data*, que permita la utilización de datos históricos para agilizar los trámites, analizar indicadores y estadísticas de los últimos años. Gracias a esta nueva inserción, se desea reinventar las estrategias en la asignación de recursos destinados a la adjudicación de créditos, a conocer las regiones donde pertenecen los usuarios, la profesión de preferencia, entre algunas otras características. En definitiva, el objetivo es tomar decisiones que impacten en los diferentes territorios del país, que los jóvenes tengan acceso a la educación superior sin tener que irse de sus ciudades natales y revisar las líneas de crédito actuales para entender cuáles son las preferencias de los jóvenes que ingresan a estudiar un programa de pregrado o un postgrado.

Una estrategia de buen uso de datos, según Fisher (2009), puede ayudar a una empresa a tomar decisiones de alto impacto, a mejorar los indicadores de áreas de la organización y lograr un trabajo de alto nivel. Por otro lado, Nasció (2008), indica que la mejor arquitectura de datos tendrá un impacto completo cuando esté alineado al negocio de la compañía.

Ahora bien, el ICETEX, durante sus años de funcionamiento, no ha establecido una estrategia basada en datos. Con esta urgencia, la alta dirección debe centrar sus esfuerzos para que dicha estrategia sea posible puesto que es ella el área encargada de planear los objetivos de cada año, los indicadores empresariales y la conformación de un equipo de expertos que logren implementar estas tecnologías.

## Capítulo 3. Marco conceptual

### 3.1. Revisión de literatura

En este capítulo, se describe el marco teórico sobre *Big Data* que soporta la propuesta de este proyecto. Por ello, se realizará una descripción de algunos conceptos relevantes, los modelos de arquitectura basada en *Big Data* y sus campos de aplicación en empresas. Este último elemento es fundamental para la investigación puesto que estas entidades se pueden presentar como casos de éxito en el análisis de datos.

*Big Data* es un enfoque práctico que requiere de diferentes transformaciones de la información y que plantea desafíos para las compañías que quieran experimentar o adoptar este nuevo activo. *Big Data* no es solo análisis de datos, también es infraestructura, competencia y, ¿por qué no?, base de las soluciones futuras, si se logra una buena implementación.

Un ejemplo de ello son las compañías que realizan ventas en línea y utilizan las bases de datos para escoger el público más conveniente que compraría sus productos o servicios específicos. Identificar y clasificar los datos es importante para las empresas porque permite ordenar las características respecto a las ciudades, las comidas, los tipos de mercado, las necesidades y las prioridades de cada grupo poblacional, divididos tanto en géneros como focos generacionales.

Oracle hace referencia al *Big Data* en los siguientes términos: “la arquitectura típica ha cambiado significativamente a lo largo de los años, debido a la gran variedad de fuentes de datos, donde se ve sistemas y arquitecturas completamente nuevas para las empresas, por ejemplo, el aprendizaje automático, minería de datos, análisis predictivo y la visualización de grandes datos” (Batra, 2015).

La gran cantidad de datos generados por las diferentes fuentes, y desde la década del 2000, se ha masificación por el uso de los dispositivos móviles, pues hicieron que los medios tradicionales de almacenamiento y procesamiento, creados en los años ochenta para la recolección selectiva de datos estructurados, resultaran insuficientes. Por lo anterior, la primera referencia al término *Big Data* consistió en la identificación del reto tecnológico que implicaba la generación de datos cuyo volumen, velocidad y variedad, desafiaba los sistemas tradicionales (Laney, 2001).

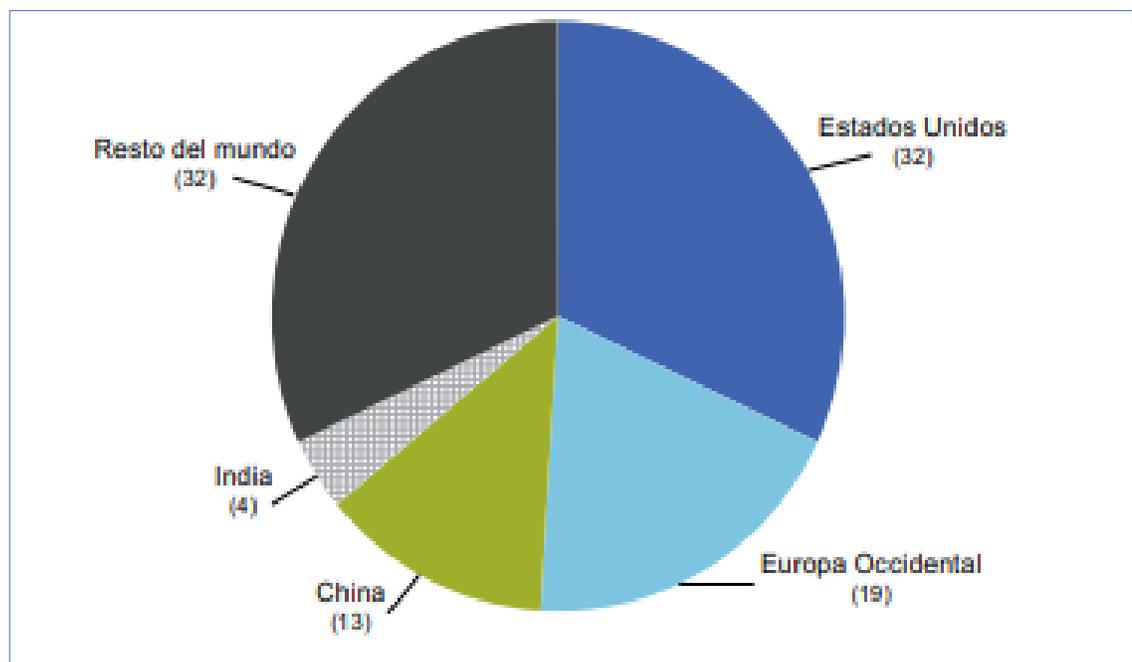
Según la revista *La nueva revolución digital* (CEPAL, 2017), el análisis de datos implica un proceso y un uso particular de estos, pues su finalidad va más allá de la capacidad de las herramientas de software especiales para bases de datos. Si bien las tecnologías relacionadas con el manejo de datos capturan, almacenan, gestionan y analizan información, estos procesos requieren de altos niveles de velocidad y expansión rápida en tiempo real. La

importancia de las tecnologías digitales permite crear estrategias o cambiar de decisiones en un instante de tiempo.

El enorme aumento de datos compensa su carácter desordenado y confuso. Si bien es cierto los datos abren alternativas para mejorar la toma de decisiones y dar soluciones de sistemas complejos, es necesario ofrecer recomendaciones y acciones que respondan a obstáculos del *Big Data*.

La Organización de las Naciones Unidas (ONU) trabaja para crear una alianza sobre los datos de desarrollo sostenible. Sin embargo, hay países que cuentan con datos mayores en comparación con otros. En la figura 5, se puede observar el número de porcentaje que corresponde a cada una de las regiones a nivel mundial. De estos porcentajes, se evidencia que, entre Estados Unidos, Europa occidental, India y la China tienen el 68% de los datos a nivel mundial.

Figura 5. Porcentaje de datos de regiones desarrolladas



Fuente: Cepal (2017).

Algunos autores resaltan que las acciones de la inteligencia de negocios, en la banca de Latinoamérica, serán más relevantes con el avance de los años, pues vienen siendo desarrolladas en forma continua por empresas de tecnología, como Google, Facebook y Amazon. Estas presentan características más recurrentes en el uso de las *Startups Fintech*, ya que tienen información de los clientes para la entrega de ofertas en tiempo real o para la optimización de los procesos en un trámite del banco. (Heredia, 2017, 99)

Por otro lado, el término inteligencia de negocios incorpora diferentes tecnologías, plataformas de software, aplicaciones informáticas y procesos para contribuir a la toma de decisiones que mejoren el desempeño de la empresa y originen una ventaja competitiva en el mercado. Los impactos de la toma de decisiones se presentarán en una mayor rapidez, en la conversión de datos a información y el uso de una aplicación para la administración. (Peña, 2006, p 22).

De igual manera la rapidez de la conversión de datos no se podría realizar si no fuera por el uso de los sistemas de información. Un sistema de información es un conjunto de elementos que interactúan entre sí, con el fin de apoyar las actividades de una empresa o negocio en las fases de entrada, almacenamiento, procesamiento y salida de información. (Peralta, 2000, p 25), estas cuatro actividades se ven reflejadas en algunos modelos o metodologías existentes para hacer análisis de datos y presentar estrategias.

Una de ellas es la entrada de información o la captura de los datos que se refiere al proceso de entrada; sus datos pueden ser obtenidos de manera manual o automática. La segunda fase es la de almacenamiento de información, donde se presenta las actividades o capacidades más importantes que tiene un sistema de información para guardar los datos mediante estructuras en archivos. La tercera fase está relacionada con el procesamiento de información, que es la capacidad que tiene el sistema de información para realizar cálculos según la sentencia que se ha dado. Por último, la salida de información, se da cuando esta se presenta por diferentes terminales como impresoras, monitores, memorias, etc.

Ahora bien, según un artículo publicado en el World Economic Forum (WEF), indicó que cada minuto de cada día, cada acción, reacción, decisión, evento y proceso se expresa a través de datos, que recopilados adecuadamente generan conocimiento. Los datos para las empresas son fundamentales para adquirir una ventaja competitiva, aun considerando que solo el 20% de estos se pueden consultar actualmente. Ello significa que hay un 80% sin procesar o son de empresas privadas. (Rometty, 2018, p 1).

Algunas empresas de tecnología, trabajan en capturar datos que están sin procesar, aproximadamente esto es un 80% y su recopilación posiblemente ayudará a desarrollar perspectivas más profundas y tomar decisiones que brinden sostenibilidad y modernización a una empresa. De igual manera, estar a la vanguardia de herramientas tecnológicas que involucran usuarios de las diferentes páginas de compañías que promocionan servicios, resulta ser motivo vital para la completa recopilación de los datos, ya que estos generan valor.

### **3.2. Arquitectura empresarial basado en *Big Data***

*Big Data* es la iniciativa para la alineación de la arquitectura empresarial. Según McKinsey: “la recompensa de unirse a la revolución *Big-Data* ya no está en duda. La investigación más amplia sugiere que cuando las empresas inyectan datos y análisis en sus operaciones, pueden ofrecer productividad y ganancias que son más altas que las de la competencia ”(Oracle, 2016, p.11)

De igual manera, *Big Data* se puede describir desde el punto de vista de un dato simple a múltiples interpretaciones según el contexto donde se investigue. Si bien existen muchas técnicas para analizar los tipos de datos, las compañías han adoptado Apache Hadoop y base de datos NoSQL (Oracle, 2016, 4). Asimismo, Oracle define que *Big Data* debe tener volumen, velocidad variedad y valor. En la tabla 6, se describen las características de estas cuatro “V”.

Tabla 6. Características determinantes para implementar *Big Data*

Característica	En qué consisten	Propósito/Recomendación
Volumen	Gran cantidad de datos de baja densidad, como datos de Twitter, clics de página web, tráfico de red	El propósito de <i>Big Data</i> es convertir estos datos que son de baja densidad en alta densidad
Velocidad	Estos datos van directamente a la memoria, y requiere uso de buena infraestructura tecnológica.	Para este caso es importante evaluar datos en tiempo real y enviar respuesta, como comercio electrónico aplicaciones de Internet de las cosas (IoT), aplicaciones móviles en tres otras con expectativa de una respuesta inmediata.
Variedad	Datos no estructurados y semiestructurados, como texto, audio y video.	Estos datos tienen mayor complejidad y de lo contrario a la velocidad se convierten una carga en tiempo real.
Valor	El desafío de esta característica es lograr identificar preferencia o sentimientos del consumidor.	Aquí se requiere a analistas inteligentes, usuarios comerciales, reconocimiento de patrones y poder predecir un comportamiento.

Fuente: Elaboración propia a partir de artículo “Arquitectura basada en *Big Data*”

Uno de los primeros casos fue un análisis de registro web minorista, mediante el cual se informó los resultados decepcionantes de los canales web durante la temporada de navidad. La intención primera de la empresa era mejorar la experiencia de los clientes en el sitio de compras en línea, pero al notar que las compras no eran las esperadas, los analistas de la compañía investigaron el patrón de navegación del sitio web, especialmente, los carritos de compras abandonados, que en este caso se tuvo en cuenta el volumen, pero debe estar acompañado de las otras variables (Oracle, 2016, 38).

Para dar respuesta a este problema, los analistas identificaron inconvenientes en cuanto, al número de desarrolladores con experiencia en SQL, pues no contaban con el número suficiente. De igual manera, con los desarrolladores y con los datos recopilados, Oracle permitió que los programadores accedieran al historial de los movimientos rechazados, y así poder realizar un procesamiento y análisis. Todo para realizar una estrategia en las ventas.

Por otro lado, *Big Data* se utiliza para aquellos grupos de bases de datos que, por su elevado volumen, diversidad y complejidad, no se pueden analizar, visualizar y almacenar con las

herramientas y almacenes de datos tradicionales. Por esto, a este concepto también se le asocia términos de "Datos masivos o "Datos a gran escala" (Chandarana y Vijayalakshmi, 2014).

En argumenta que son muchas las características o definiciones que existen de *Big Data*, no obstante, la mejor forma de describir esta tecnología es a partir de parámetros específicos denominados las V de *Big Data*. Estas son: Volumen, Velocidad, Variedad, Veracidad, Valor, Validez y Visualización. Gracias a estos estudios las diferentes empresas pueden analizar una estrategia para implementar una arquitectura empresarial basada en *Big Data*.

### 3.4 Proceso de desarrollo de Oracle de arquitectura empresarial

Oracle, en su artículo *An Enterprise Architect's Guide to Big Data (2016)*, indica que existen buenas prácticas en el uso del *Big Data* y que para obtener los máximos beneficios de este es necesario no comprenderlo como una isla. *Big Data* es un universo de información a niveles desorbitantes, que describe una gestión que integra diferentes tipos de datos. Estos datos, como se mencionó en la tabla 6, tienen características de valor, variedad y volumen que se deben tratar en la arquitectura de información y alinearla al negocio de la empresa y según su característica al área en específica.

Del mismo modo, el artículo afirma que las organizaciones están asumiendo cada vez más riesgos y experimentando con *Big Data*. Por esta razón, el concepto se debe entender como un activo empresarial para alinearlo con la estrategia central de la compañía.

*Big Data* presenta nuevas formas de hacer las cosas en un entorno empresarial y describe seis pasos para establecer un contexto y el alcance del negocio.

- Paso 1: establecer el contexto y alcance del negocio

En el contexto del negocio es importante conocer a qué se dedica la compañía, en este paso las organizaciones no están seguras de lo que posiblemente pueden enfrentar y tienen un desafío para definir un alcance y es muy importante tener el personal capacitado como profesionales y arquitectos en *Big Data* que pueda identificar y tener una visión general del valor y el retorno de la inversión.

En este caso, el ICETEX otorga créditos a personas que se gradúan de educación media y pretenden continuar con estudios en educación superior. Generalmente, estos aspirantes son de estratos 1, 2 y 3, que por algún motivo económico deben acudir a créditos para pagar sus estudios.

Tabla 7. Fase 1 Contexto y alcance del negocio

Usos	preguntas	Enfoque General	Enfoque ICETEX
Intento comercial	¿Cómo hacer uso de los datos?	Vender nuevos productos y servicios. Personalizar experiencias de clientes. Detecta las necesidades de mantenimiento del producto. Predecir riesgos, resultados operativos. Vender datos de valor agregado	<ul style="list-style-type: none"> <li>Mejorar las líneas de crédito, para facilitar el acceso al crédito educativo y demás productos del ICETEX.</li> <li>Implementar estrategias de comunicación para mejorar la imagen de la entidad.</li> </ul>
Uso comercial	¿Qué procesos comerciales pueden beneficiarse?	Sistemas operativos ERP / CRM BI y sistemas de informes Análisis predictivo, modelado, minería de datos	<ul style="list-style-type: none"> <li>Atención al cliente.</li> <li>Integración de sistemas y uso de nuevas herramientas</li> </ul>
Propiedad de datos	¿Necesitamos archivar los datos?	Propietario Requerir datos históricos Asegurar el linaje Gobierno	Organizar los datos existentes en las bases de datos del ICETEX para un mejor uso.

Fuente: Elaboración propia adaptada de arquitectura empresarial Oracle

- Paso 2: establecer una visión de arquitectura

Cumplir con el alcance del negocio es crear una “gran idea”, que identifique las Fuente:s de los datos, desarrolle y mejore las hipótesis para interpretar y refinar los análisis y, así explorar los resultados.

Los datos del ICETEX se podrían analizar en todos los sectores ya que los usuarios estudian programas académicos que pueden contribuir a las diferentes áreas de la industria y necesidades que requiere el país que aplicaría a una gran idea empresarial.

Tabla 8. Fase 2. Establecer una versión de arquitectura

Usos	preguntas	Enfoque General	Enfoque ICETEX
Ingestión	¿Cuáles son las características de sentido y respuesta?	Eventos en tiempo real basados en sensores Eventos de transacciones casi en tiempo real Análisis en tiempo real Análisis casi en tiempo real	<ul style="list-style-type: none"> <li>Aprobación de créditos en tiempos cortos.</li> <li>Respuestas y soluciones de PQR en línea.</li> </ul>

		Sin análisis inmediatos	
Almacenamiento de datos	¿Qué tecnologías de almacenamiento son mejores para nuestros datos? ¿reservorio?	HDFS (Hadoop más otros) Sistema de archivos Almacén de datos RDBMS Base de datos NoSQL	<ul style="list-style-type: none"> <li>• BD NoSQL ya que ayuda a la flexibilidad de almacenar, ordenar y capturar datos no estructurados y no estructurados.</li> <li>• Almacén de datos.</li> <li>• Bases de datos relacionales en la nube.</li> </ul>
Procesamiento de datos	¿Qué estrategia es práctica para mi aplicación?	Déjalo en el punto de captura Agregar transformaciones menores Datos ETL a plataforma analítica Exportar datos a escritorios	<ul style="list-style-type: none"> <li>• Analítica desde archivos sencillos a los complejos.</li> <li>• Facilidad para exportar datos.</li> </ul>
Actuación	¿Cómo maximizar la velocidad de consulta de transformaciones de datos y modelado analítico?	Analizar y transformar datos en tiempo real Optimizar las estructuras de datos para el uso previsto Usar procesamiento paralelo Aumentar hardware y memoria Configuración y operaciones de bases de datos Dedicar cajas de arena de hardware Análisis de datos en reposo	<ul style="list-style-type: none"> <li>• Optimizar las estructuras de los datos.</li> <li>• Aumentar hardware y memoria.</li> <li>• Configuración en las bases de datos.</li> </ul>
Latencia	¿Cómo minimizar la latencia entre operaciones clave y componentes? (ingesta, depósito, almacén de datos, informes, sandboxes)	Compartir almacenamiento Interconexión de alta velocidad Red privada compartida VPN: en redes públicas	<ul style="list-style-type: none"> <li>• Compartir almacenamiento.</li> <li>• Interconexión de alta velocidad.</li> <li>• Red privada compartida.</li> <li>• VPN: en redes públicas.</li> </ul>
Análisis y descubrimiento	¿Dónde necesitamos hacer análisis?	Al ingerir - evaluación en tiempo real	<ul style="list-style-type: none"> <li>• En los datos de las líneas de crédito.</li> </ul>

		En un depósito de datos sin procesar En un laboratorio de descubrimiento En un data warehouse / mart En herramientas de informes de BI En la nube pública En BD locales	<ul style="list-style-type: none"> <li>• En las redes sociales.</li> <li>• En las PQR que radican los usuarios.</li> <li>• Utilización de herramientas de analítica.</li> </ul>
Seguridad	¿Dónde necesitamos proteger los datos?	En memoria Redes Depósito de datos Almacén de datos Acceso a través de herramientas y laboratorio de descubrimiento	<ul style="list-style-type: none"> <li>• En las BD que se tienen</li> <li>• En los repositorios</li> <li>• En las aplicaciones</li> <li>• En las redes</li> </ul>

Fuente: Elaboración propia adaptada de arquitectura empresarial Oracle

- Paso 3: evaluar el estado actual

Evaluar el estado actual de los datos implica que la compañía vea el cómo se hacen las cosas y revise las habilidades y las herramientas con las que trabaja.

Esta mirada panorámica, le dará a la entidad un diagnóstico de los activos corrientes, los modelos de datos y patrones de flujos de datos. Asimismo, se debe revisar la arquitectura tecnológica, incluidas las plataformas e infraestructura, pues así se puede indicar si los almacenes de datos tradicionales permiten que se incluya procesos para llegar a un modelo organizado de datos que generen patrones para implementar tecnologías de Big Data. En este contexto el ICETEX tiene modelo de datos en diferentes fuentes, pero si posee infraestructura que podría ser útil en este paso.

*Tabla 9. Fase 3 Evaluar estado actual*

Usos	preguntas	Enfoque General	Enfoque ICETEX
Experiencia de datos no estructurados	¿En la actualidad se están procesando datos no estructurados o del sensor en de alguna manera ? (por ejemplo, texto, espacial, audio, video)	Proyectos departamentales Dispositivos móviles Diagnóstico de la máquina Captura de datos en la nube pública Varios archivos de registro de sistemas	<ul style="list-style-type: none"> <li>• Se tiene proyectado analizar las redes sociales.</li> <li>• Analizar los sentimientos de los mensajes por medio de las</li> </ul>

			nuevas herramientas.
Experiencia de código abierto	¿Qué experiencia tenemos en Apache de código abierto? (Hadoop, NoSQL, etc.)	Experimentos dispersos Prueba de conceptos Experiencia en producción Colaborador	<ul style="list-style-type: none"> <li>• Generar una política de uso de Software Libre.</li> </ul>
Habilidades analíticas	¿En qué medida empleamos científicos de datos y Analistas familiarizados con avanzados y predictivos? herramientas y técnicas analíticas?	Si No	Crear un grupo de expertos para la implementación de <i>Big Data</i> .

Fuente: Elaboración propia adaptada de arquitectura empresarial Oracle

- Paso 4: establecer el estado futuro

Poder demostrar como a futuro se puede entregar valor comercial y, posiblemente, asegurar el éxito de la implementación de la *Big Data*. Para ello, es necesario pensar en planes de inversión ya que se requiere de infraestructura que soporte el volumen de los datos; esto podría darse en físico o en soluciones en la nube. Luego, se evaluará el impacto de las plataformas, los problemas la disponibilidad y regulaciones de seguridad que estos abarcan.

Si bien es cierto que el ICETEX puede tener infraestructura hace falta mucho, y si se debe pensar en inversión una vez sea evaluado el impacto para las posibles soluciones.

Tabla 10. Fase 4. Establecer estado futuro

áreas	preguntas	Enfoque General	Enfoque ICETEX
Mejores prácticas	¿Cuáles son los mejores recursos para guiar las decisiones a construir mi futuro estado?	Arquitectura de referencia Patrones de desarrollo Procesos operacionales Estructuras de gobierno y políticas Conferencias y comunidades de interés.	<ul style="list-style-type: none"> <li>• Lograr implementar patrones de desarrollos.</li> <li>• Estructurar mejores prácticas de gobierno y políticas para el ICETEX</li> </ul>

		Mejores prácticas de proveedores	
Tipos de datos	¿Cuánta transformación se requiere para datos no estructurados en el depósito de datos?	Ninguno Derive una comprensión fundamental con esquema o pares clave-valor Enriquecer datos	<ul style="list-style-type: none"> <li>• Enriquecer datos a través de la implementación de metodologías de analítica.</li> </ul>
Calidad de datos	¿Cuándo aplicar transformaciones?	En la red En el almacén de datos Por el usuario en el punto de uso En tiempo de ejecución	<ul style="list-style-type: none"> <li>• Mejorar el tiempo de ejecución para las respuestas de los usuarios.</li> <li>• En los resultados de los análisis de datos no estructurados como las redes sociales en tiempo real.</li> </ul>

Fuente: Elaboración propia adaptada de arquitectura empresarial Oracle

- Paso 5: desarrollar una hoja de ruta

En este punto es importante conocer el estado actual y crear una ruta para evolucionar en una alineación basada en datos y asegurar un éxito temprano de un proyecto con tecnologías que se adapten a la actualidad.

Como se mencionó en párrafos anteriores, el ICETEX no tiene un alineamiento basado en datos, pero sí puede pensar en un futuro organizado y analizado con estos datos. Los posibles beneficios que puede traer estos análisis, mejoraran el área de tecnología y capacitar al personal en estas herramientas.

*Tabla 11. Fase 5. Desarrollar hoja de ruta*

áreas	preguntas	Enfoque General	Enfoque ICETEX
POC- Prueba de concepto	¿Qué debe validar el POC antes de mudarnos? ¿adelante?	Caso de uso comercial Comprensión de nuevas tecnologías Integración empresarial Implicaciones operacionales	<ul style="list-style-type: none"> <li>• Caso de uso comercial.</li> <li>• Comprensión de nuevas tecnologías.</li> <li>• Integración empresarial.</li> <li>• Implicaciones operacionales.</li> </ul>

Habilidades de código abierto	¿Cómo adquirir habilidades de código abierto?	Entrenamiento de empleados Contratar experiencia Utilice proveedores / socios experimentados	<ul style="list-style-type: none"> <li>• Crear e implementar políticas que permitan el uso de código abierto.</li> <li>• Capacitar empleados y contratar expertos.</li> <li>• Crear competencias en línea.</li> </ul>
Habilidades analíticas	¿Cómo adquirir habilidades analíticas?	Entrenamiento de empleados Contratar experiencia Utilice proveedores / socios experimentados	<ul style="list-style-type: none"> <li>• Capacitaciones y sensibilización empresarial.</li> <li>• Adquirir herramientas que estudien analítica.</li> </ul>

Fuente: Elaboración propia adaptada de arquitectura empresarial Oracle

- Paso 6: establecer gobernanza sobre la arquitectura

Establecer una gobernanza se centra para los que tienen el acceso a los datos ya que ellos con su experiencia podrían determinar un enfoque y una adaptación de los posibles usos y, asimismo, capacitar y crear una cultura en la organización o entidad.

Lograr que el ICETEX tenga datos de calidad y permita el acceso a los mismos para realizar análisis con precisión, permitirá que se genere información muy útil para la toma de decisiones que requieren intervención e interpretación por medio de la automatización. Por dicho motivo, el enfoque de establecer gobernanza de datos debe ser rentable y adaptable en una compañía.

Oracle, en la parte técnica, presenta *Big Data SQL* como un acceso flexible en la toma de decisiones de los datos, movimiento de datos, transformación de datos, el análisis de datos, y en heredar funciones analíticas SQL. En este caso, sería conveniente que ICETEX utilice la estrategia de Oracle, dado que las bases de datos de la entidad están sobre este lenguaje.

Tabla 12. Fase 6. Establecer gobernanza sobre la arquitectura

áreas	preguntas	Enfoque General	Enfoque ICETEX
fuentes de datos en la nube	¿Cómo garantizar la confianza de las Fuentes de datos en la nube?	Administrar directamente Auditoría Asumir	Con auditoria y administrar directamente
Calidad de datos	¿Cómo limpiar, enriquecer, deducir datos no estructurados?	Usar muestreo estadístico Técnicas normales	<ul style="list-style-type: none"> <li>• Usar muestreo estadístico.</li> <li>• Técnicas vigentes que utilicen buenas prácticas.</li> </ul>

Calidad de datos	¿Con qué frecuencia necesitamos volver a validar el contenido? ¿estructura?	Sobre cada recibo Periódicamente Manual o automáticamente	Se deben validar frecuentemente y realizar actualización de los procesos manuales, pero ideal a que sea automático.
Políticas de seguridad	¿Cómo ampliar las políticas de seguridad de datos empresariales?	Heredar políticas empresariales Copiar políticas empresariales Autorizar solo herramientas / puntos de acceso específicos Limitado a monitorear registros de seguridad	Implementar políticas empresariales y monitorear registros de seguridad

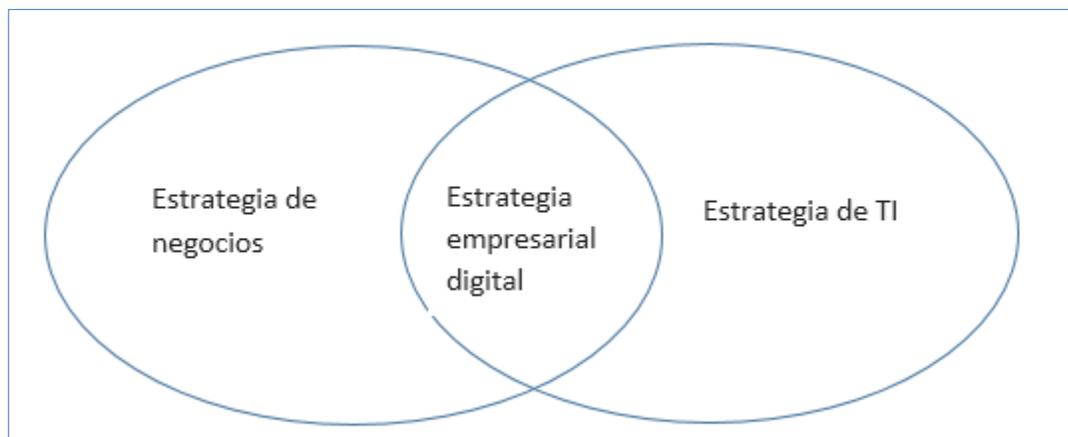
Fuente: Elaboración propia adaptada de arquitectura empresarial Oracle

### 3.5. SAVI-BIGD Marco estratégico

Lakoju (2017) en su trabajo titulado *SAVI-BIGD (Strategic Approach Of Value Identification Big Data Framework)* sugiere apropiarse y defender la implementación de una estrategia en *Big Data* para economizar costos y organizar los objetivos estratégicos de la compañía. En su investigación, identifica cinco fases: visión estratégica, implementación mapa de ruta *Big Data*, generación de objetivos estratégicos, determinación de fuentes de datos e implementación de *Big Data*.

Lakoju indica que su estrategia se implementó en dos compañías que identificaron el valor potencial de proyecto en *Big Data* y el ahorro de costos. Así mismo, menciona la fusión entre el negocio y la tecnología, la estrategia empresarial digital, como se puede observar en la figura 6.

Figura 6. Fusión estrategia empresarial y TI

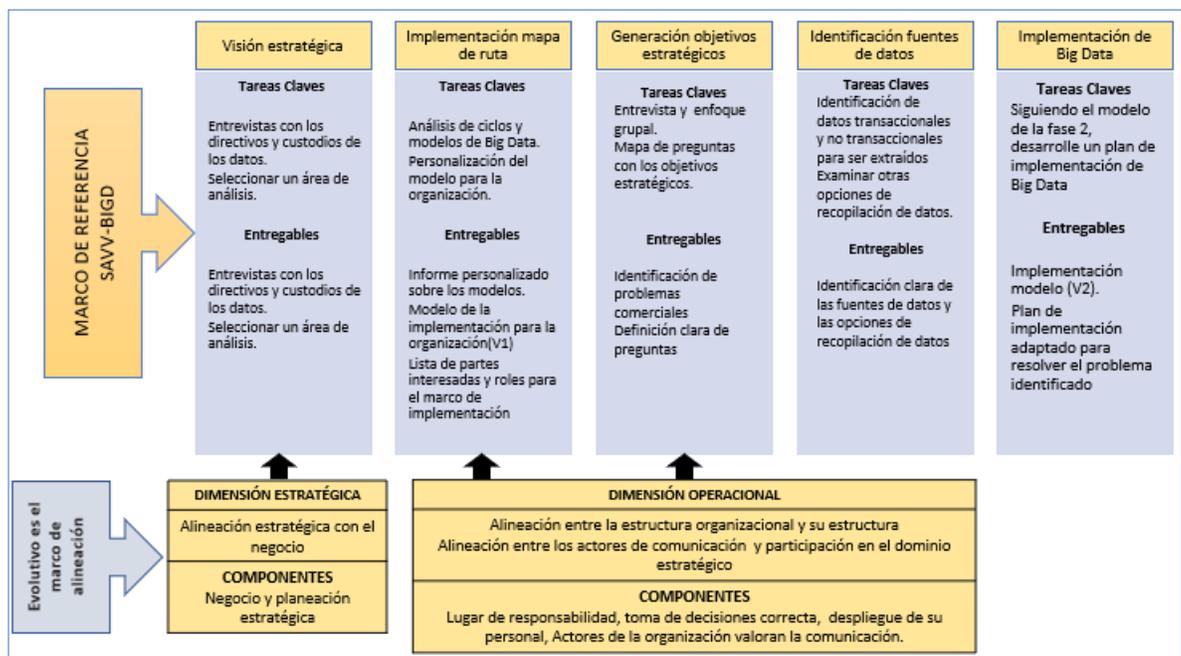


Fuente: Elaboración propia a partir de *A Strategic Approach*

*Big Data* trae información de sesión de datos, minería de datos, analítica de datos, etc. Por ejemplo, Brinkhues y Freitas (2015) argumentan que la capacidad de gestión de la información puede afectar negativamente la expectativa del costo, pero positivamente a la expectativa del valor. Por otro lado, Kung y Col. (2015) propusieron que la capacidad de las tecnologías de la información, la competencia de *Big Data*, la gestión de datos y la capacidad de organización están conectados para formar una red de factores importantes para poder decidir en algún aspecto con buena calidad.

Aunque los diferentes autores contribuyen en la literatura de *Big Data*, este marco quiere realizar el enfoque en el lado estratégico de *Big Data* para alinearlo a la estrategia de negocio y puedan sacar provecho de este.

Figura 7. Estrategia fases SAVI-BIGD marco estratégico



Fuente: Adaptada de Lakoju y Serrano (2016)

Dentro de las cinco fases que propone el SAVI-BIGD Strategy Framework se puede identificar lo siguiente:

- Fase 1: visión estratégica: implica alinear con cuidado el negocio de una compañía con las áreas de sistemas de información a través de la planificación estratégica. Las áreas o altas gerencias de una compañía se deben alinear y tener en cuenta el área de tecnología, pues ayudará a direccionar la solución de un problema por medio de los datos.

- Fase 2: Diseño de hoja de ruta para *Big Data*: en esta fase, la estructura de la organización y los sistemas de información interactúan con la estrategia. Se identifican y se alinean a las prioridades del área comercial para solucionar los problemas identificados por medio de una implementación de *Big Data* adecuada.
- Fase 3: Generación de objetivos estratégicos de Big Data: para esta fase, también se debe alinear la estructura organizacional y los sistemas de información de acuerdo con los problemas comerciales de la compañía. Mientras se evalúa el problema, mediante un estudio cualitativo con las partes interesadas, el problema puede ser analizado a través de métodos de recolección de información y de ciertos análisis con los datos que cuenta la entidad.
- Fase 4: Identificación de fuentes de datos: una vez analizadas las anteriores fases, se identifica las posibles fuentes de datos a analizar para solucionar un posible problema e identificar una estrategia que impacte el negocio.
- Fase 5: Implementación de *Big Data*: en esta fase, se tiene en cuenta la infraestructura y las necesidades de los usuarios, analizadas en las anteriores fases. Para ello, se buscará un instrumento o una herramienta que permita identificar los criterios para la solución de los problemas que se encontraron, priorizando por necesidad.

### 3.3 Modelos de negocios basados en datos

La Universidad de Cambridge presenta un estudio con la intención de invitar a las organizaciones a implementar un modelo de negocio basado en datos. En él, se sugiere que un modelo de negocio debe ser diseñado para crear valor comercial y ganancias que les den sostenimiento a las empresas. De esta manera, la Universidad de Cambridge, en su artículo “Data and Analytics - Data-Driven Business Models: A Blueprint for Innovation” (2015), presenta seis inquietudes para que las empresas se planteen la urgencia de *Big Data* y puedan generar un modelo de innovación de acuerdo con los datos existentes.

Para ello, se analizará cada una de las preguntas que presentan los autores, se abordará el vacío que tienen las empresas y su posible beneficio para implementar un modelo basado en datos

Pregunta 1. ¿Qué se quiere lograr usando *Big Data*?

En este caso, las organizaciones deben tener definidos sus objetivos estratégicos; estos deben ser claros, realistas y alcanzables apuntando a un resultado potencial asociado con los datos y con los recursos que cuenta la organización a nivel de talento humano como de infraestructura. En las empresas analizadas, según el artículo de Cambridge, se encontraron siete ventajas: cadena de suministro acortada, expansión, consolidación, velocidad de procesamiento, diferenciación y marca.

Se considera que la marca es la ventaja competitiva más importante para organizaciones establecidas, pues el 95 % de las empresas analizadas lo consideraron como una ventaja competitiva, seguido por la diferenciación (90 %) y la expansión (70 %). Por su parte, la cadena y la velocidad de procesamiento se consideraron menos significativas para las organizaciones que se analizaron. En este contexto el ICETEX quiere lograr identificar las diferentes demandas educativas, preferencias de programas académicos, y por qué no pensar en las necesidades que tiene el país por falta de personal capacitado para los diferentes empleos que están vacantes, esto como una ventaja competitiva.

Pregunta 2. ¿Cuál es la oferta deseada?

Brownlon, Zaki, Neely y Urmetzer afirman que las empresas tienden a utilizar los datos para mejorar o aumentar su oferta, se piensa en la propuesta valor para los clientes. En un primer momento, se piensa en los clientes para reconocer sus necesidades y responder a ellas; la segunda, respectivamente, para crear las mejores ofertas de valor de acuerdo con la segmentación y el tipo de clientes que tiene la organización.

Por lo anterior, se pueden referir algunos de los objetivos del ICETEX como, por ejemplo, otorgar créditos para estudio que les permitan a los usuarios iniciar, permanecer y terminar su programa académico, son clientes claramente identificados, y se pueden definir:

Consumidores individuales (B2C): son todos los estudiantes que de alguna manera tienen un producto que con un buen análisis se pueden priorizar por algún patrón determinado, por otro lado estrictamente los clientes en empresas (B2B): para las diferentes entidades, constituyentes que requieren una administración de un porcentaje de recursos para otorgar créditos, subsidios o becas a personas que quieran realizar algún tipo de estudio y por último el consumidor a consumidor (C2C): Este modelo se puede utilizar tanto en los clientes nuevos, que pueden llegar a ser un consumidor o un tercero que quiera invertir en la administración de unos recursos para la educación

Pregunta 3. ¿Qué datos se requieren y cómo van a ser adquiridos?

De acuerdo con el tipo de negocio de la compañía, se debe estudiar cuáles son los datos que se van a adquirir. Para ello, será necesario que la empresa opte por la compra de datos que algunas otras compañías tienen o por la revisión de datos ya existentes en sus bases de datos, pero que aún no han sido analizados ni recopilados. Según la investigación de la universidad de Cambridge, los datos proporcionados por los clientes fueron utilizados y analizados en un 80%, con datos existentes. Estos análisis tienden a orientar a las empresas a convertirse en estrategias impulsadas por datos.

Por ejemplo, las minoristas de moda analizaron datos gratuitos y combinados con los proporcionados por los clientes ejecutaron predicción y protocolo de análisis descriptivos para determinar las tendencias dentro de este comercio. De igual manera para el ICETEX se pueden adquirir los datos suministrados por los usuarios de algún producto o crédito activo, como se ha mencionado son aproximadamente 744.000 estudiantes activos el cual tienen un considerable número de datos; datos de uso colaborativo a los que se les puede aplicar diferentes tipos de análisis.

Pregunta 4. ¿De qué manera se van a procesar y aplicar esos datos?

Para el procesamiento de datos, se debe realizar una clasificación de datos, un rastreo a las posibles actividades claves del negocio y su impacto en la planificación, una revisión a las necesidades de las aplicaciones de software y los recursos de hardware; todo esto en relación con los colaboradores que tengan habilidades o experiencia en herramientas que analicen gran cantidad de datos. La clasificación de datos se puede realizar de forma manual, por medio de los colaboradores mediante alguna herramienta que se tenga instalada y licenciada para realizar de una forma programada; o automática con una herramienta específica para un análisis especializado. Como se mencionó anteriormente, el análisis genera información que se pueden subdividir en:

- Análisis descriptivo, actividades analíticas que explican el pasado;
- Análisis predictivo, que predecir / pronosticar el resultado futuro;
- Análisis prescriptivos, que predicen resultados futuros y sugerir decisiones

Por lo anterior, el ICETEX puede revisar cuál es la mejor opción para aplicar según la necesidad específica de cada área, desde tecnología se podrá orientar de un alineamiento o una estrategia basada en la aplicación de datos.

Pregunta 5. ¿Cómo se va a monetizar?

Es importante tener claro que implementar un negocio basado en datos va a generar un beneficio cuantificable para el éxito operativo y posicionamiento. Hartmann, Zaki, Feldmann y Neely identifican Fuente:s de ingresos como la venta de activos, regalo a la propiedad de derechos de un bien o servicio a cambio de dinero; préstamo, alquiler, leasing, concesión temporal, derecho exclusivo de un activo por tiempo definitivo, licenciamiento, otorgamiento de permiso para usar una propiedad, entre otros (2014, 10). Estos modelos difieren considerablemente a una tarifa de suscripción.

Si se toma como ejemplo el ICETEX, vale la pena afirmar que, dados los objetivos del plan estratégico de la entidad, "contribuir al propósito nacional de “hacer de Colombia la mejor educada” elevando la calidad de vida y contribuyendo a la prosperidad del país”, estas estrategias se pueden presentar como un beneficio cuantificable y rentable para la entidad y para la empleabilidad del país.

Pregunta 6. ¿Cuáles son las barreras para lograr el objetivo?

En el análisis de la investigación de la universidad de Cambridge, se indica que las personas pueden llegar a convertirse en una barrera al no recibir los cambios tecnológicos con facilidad. Por ello, el problema de personal se convierte en un asunto cultural, y no solo por ser una entidad pública que está regida por temas de contratación, licitaciones y demás normas que no se pueden ignorar.

Frente a la falta de personal especializado en análisis de datos, se puede indicar la reducción de innovación en la implementación de posibles estrategias basadas en datos. Actualmente se habla que los datos son el nuevo “oro” (Deloitte, 2020) que impactan decisivamente en el futuro de las empresas que prestan servicios. Para involucrar al ICETEX en estas posibles barreras se debe involucrar a todo el personal, realizando capacitaciones que cambien el pensamiento empresarial y puedan implementar las mejores prácticas del análisis o estudio de los datos y así llegar a la utilización del *Big Data*.

### 3.6 Guía Técnica MinTIC – Gobierno de Datos

La guía técnica del Ministerio de Tecnologías de la Información y las Comunicaciones - MinTIC presenta una metodología práctica e indica los posibles aportes de una arquitectura del dato, como: el valor que puede representar en las entidades del Estado, la coherencia en los conceptos y datos manejados, los tiempos de puesta en producción y el control de datos sensibles. Por otro lado, destaca el valor que puede generar la reutilización de componentes de manejadores de repositorios y bases de datos tales como servicios e interfaces, pues implica reducción de esfuerzos y menores tiempos de implementación, asegurar la coherencia de los datos desde el mismo diseño de la solución y la estandarización y homogeneización en la modelación de datos (MinTIC, 2014).

#### 3.6.1 Descripción

Dentro de la guía de MinTIC describe tres procesos para una transformación:

Eficiencia operativa (valor para TI): flujos de datos más eficientes, menos incidencias y rechazos por calidad de datos; trazabilidad, reutilización de Componentes, robustez de las soluciones y rendimiento; conocimiento del dato (valor para la función pública): visión integrada y unificada 360° del ciudadano y grupos de interés, campañas mejor orientadas, convergencia, control de datos sensibles y coherencia de datos; uso de información (valor en la transformación): se refiere al uso de información de calidad en los procesos o iniciativas de migración / fusión de datos hacía una institución convergente, multicanal y con una oferta simplificada. Facilitando con esto la gestión de la transformación y el proceso de migración de datos.

#### 3.6.2 Beneficios

Para identificar los beneficios es importante que las entidades coordinen la estructura y los proceso de:

*Figura 8. Importancia del dato*



Fuente: Elaboración propia.

Este ámbito es muy grande y las necesidades de cada compañía son diferentes por tal motivo se debe identificar el nivel de madurez para la elección de un marco que determine cada entidad o empresa.

### 3.6.3 Flujo de pasos del gobierno de datos.

Cuando una compañía o una institución decide implementar el gobierno de datos es importante considerar los siguientes pasos:

Figura 9. Pasos gobierno de datos



Fuente: Elaboración propia.

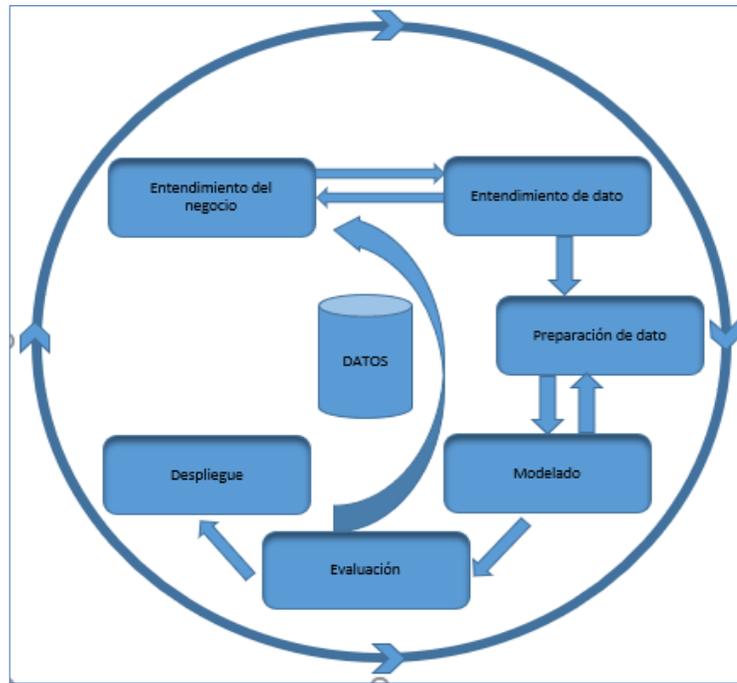
Estos pasos se deben trabajar de una forma transversal y continua, en fin, de permitir auditoría y mejoras para el dato, por otro lado, esta estrategia utiliza datos maestros, metadatos funcionales, responsabilidades de gobierno, mecanismos de acceso, datos orquestados entre sistemas y normativas para codificación y reglamentación del ciclo de vida del dato.

### 3.7. Metodologías de analítica basados en *Big Data* – Metodología CRISP-DM

La metodología CRISP-DM describe un estándar que incluye un modelo y una guía, estructurada en seis fases. Algunas de ellas permiten revisar parcialmente un paso o la totalidad de la fase terminada. El modelo pretende construir variables que sirvan como Fuente: de información, siguiendo pasos que ayuden a comprender el mejor camino para la toma de una decisión, generación de un nuevo conocimiento e impacto en el negocio de la empresa.

En la figura 10, se presenta el ciclo de la metodología Crisp-DM

Figura 10. Ciclo de metodología CRISP-DM



Fuente: (Nadali, 2011)

En la figura anterior se observan las fases del modelo CRISP-DM donde se evidencia que algunas son bidireccionales. En seguida se describe cada una de las fases.

### 3.7.1 Fase I. Comprensión del negocio

En esta fase se debe reconocer los objetivos y los requerimientos de un proyecto. A su vez, se realiza el conocimiento de los datos para definir un problema y un plan para lograr los objetivos.

### 3.7.2 Fase II. Comprensión de datos

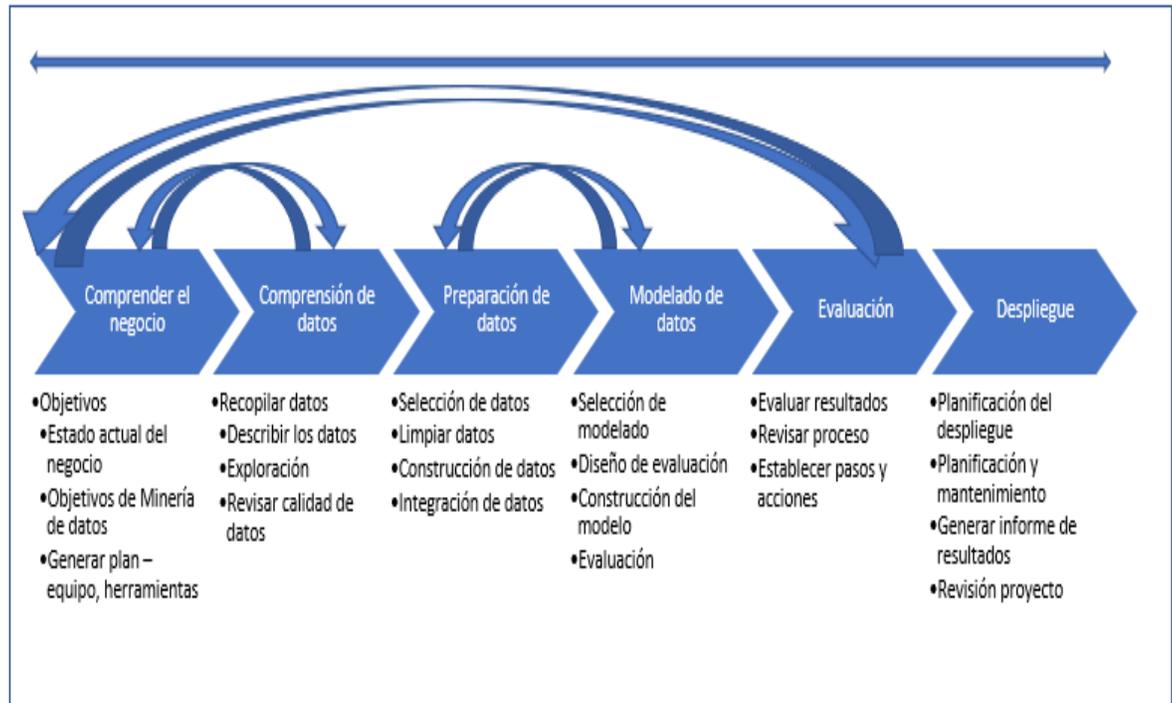
Es la colección de datos que se tiene para luego realizar una serie de actividades que ayudan al entendimiento de estos. A partir de estos, se identifica si tienen problemas de calidad, ayudando a generar un conjunto preliminar y posibles subconjuntos que ayudan a formar premisas en cuanto a la información que aún no se ha podido llegar.

### 3.7.3 Fase III. Análisis de datos y selección de características

Esta fase comprende todas las actividades necesarias para construir el conjunto de datos que se van a utilizar en la herramienta que se tenga para el modelado. Entre sus procesos se

requiere una selección de tablas, registros y atributos, haciendo la limpieza de los datos para modelarlos.

*Figura 11. Ciclo de metodología.*



Fuente: elaboración propia

Para las siguientes tres fases, se pretende tener un grupo de expertos, entre ellos: arquitecto de datos, científico de datos, analista de datos y un especialista en las diferentes herramientas que puedan facilitar el trabajo del ciclo de esta metodología.

#### 3.7.4 Fase IV. Modelado

En esta fase se aplican técnicas de modelado que involucren el problema, ajustando parámetros para tener valores óptimos. Existen diferentes formas de dar solución a un tipo de problema de minería de datos. Sin embargo, sea la solución que se dé, se deben preparar los datos para realizar el proceso.

#### 3.7.5 Fase V. Evaluación (obtención de resultados)

Una vez se ha realizado el análisis junto con las otras fases, se procede con una evaluación que consiste en revisar los pasos ejecutados para crear y comparar el modelo obtenido con los objetivos del negocio. Es importante revisar que no haya quedado ningún interrogante sobre el negocio porque al final de esta fase se debe tener una decisión sobre la aplicación de los resultados en el análisis de los datos.

### 3.7.6 Fase VI. Despliegue (puesta en producción)

Luego de realizar la creación del modelo, ahora sí con el conocimiento de los datos, se pueden presentar los datos al cliente para que pueda usarlos. Una vez se realice el desarrollo del modelo se evidencia la complejidad o se puede pensar en automatizar el proceso de análisis de los datos.

A continuación, se referencian algunos casos que incluyeron el uso de *Big Data* siendo exitosos y generaron políticas para fortalecer la proyección del país.

## 3.8. *Big Data*: casos de estudio aplicables en el contexto

### 3.8.1 Consejo Nacional de Política Económica y Social -CONPES 3920

De acuerdo con el Decreto 1743 de 2016, el Sistema Estadístico Nacional - SEN es definido como “el conjunto articulado de componentes que, de manera organizada y sistemática, garantiza la difusión de las estadísticas oficiales que requiere el país”. En este sentido el SEN está compuesto por las entidades que producen y difunden estadísticas o son responsables de registros administrativos. Por lo anterior y según el artículo 160 del Plan nacional de Desarrollo-PND 2014-2018 *Todos por un nuevo país*, indico que el SEN realizaría estadísticas oficiales a partir del aprovechamiento de registros administrativos y del intercambio de microdatos entre las entidades del SEN con altos estándares de calidad y de manera oportuna. De igual forma se definió al DANE como rector del SEN. Por lo anterior es importante resaltar que en el PND 2014-2018, determina expresamente la necesidad de disponer de una política pública de explotación de datos en un sistema de gestión de calidad.

De esta manera el Departamento Nacional de Planeación y el Ministerio de Tecnologías de la Información y las Comunicaciones liderarán la implementación de la Política nacional de explotación de datos (Big Data), en un periodo de implementación de cinco años, con inversiones por 16.728 millones de pesos para la esta inversión.(CONPES, 2018)

Por otro lado, las prioridades de algunos países se agrupan en: aumento de innovación y capacidades científicas (Estados Unidos), mejorar la eficiencia administrativa (Australia), fortalecer las capacidades industriales (Francia, Corea, Japón, China) y promover el crecimiento económico en el contexto de la economía digital (Reino Unido, Unión Europea). Adicionalmente, las políticas de estos países resaltan la necesidad de anticiparse a los riesgos potenciales que pueden surgir de usos indebidos del aprovechamiento de datos, principalmente: intromisiones en la privacidad y la intimidad, existencia de monopolios, prácticas discriminatorias y limitaciones a la libertad de elección y de acceso a la información (Greenwood, Stopczynski, Sweatt, Hardjono, & Pentland, 2014).

MinTIC tiene como presupuestado para el 2021 incluir acciones de apropiación de TIC, que son proyectos para mejorar la calidad de vida de los colombianos reduciendo la brecha digital y fomentar el uso de la explotación de datos, así como la apropiación de los productos que esta genera, orientado a dinamizar el ecosistema de datos en el país. Lo anterior permitirá que los ciudadanos incorporen en su vida diaria los beneficios que pueden obtener de los datos y de los bienes, servicios y productos basados en estos, así como tener la capacidad de encontrarlos, determinar su credibilidad, producirlos e interpretarlos de manera crítica.(Conpes, 2018, 89).

### 3.8.2 Aplicación de *Big Data* en la educación, minería de datos y aprendizaje de análisis.

Según el artículo de Katrina Sin, Loganathan Muthu (2015). La investigación en educación ha dado como resultado mejoras pedagógicas. En el aprendizaje actual, los usuarios aprenden en discusión de foros, chats en línea, clientes de mensajería instantánea y varios sistemas de gestión de aprendizaje como Moodle, aprendizaje en línea para mejorar la experiencia. Por otro lado, han implementado entrada de proyectos de código abierto en dispositivos móviles. Los estudiantes han comenzado a usar teléfonos inteligentes para acceder al contenido de aprendizaje. Estos entornos de aprendizaje tienen hacerse accesible en cualquier lugar a través de internet, los estudiantes acceder a sus cursos en cualquier lugar y disfrutar de actividades de aprendizaje. Las actividades de los estudiantes a través de sistemas de gestión de aprendizaje crean gran cantidad de datos que pueden utilizarse para desarrollar nuevas alternativas y mejoras según sus experiencias, además estos datos se vuelven disponibles para la utilización de cursos, clases y estudiantes. Gracias a estos datos y las limitaciones convencionales las instituciones educativas iniciaron a explorar tecnologías de Big Data, para procesar datos educativos. En las aplicaciones de aprendizaje se pueden mencionar:

- Predicción de rendimiento: se puede analizar con la interacción del estudiante en un entorno de aprendizaje con otros estudiantes y profesores
- Detección de riesgo de desgaste: este se puede determinar cuando el comportamiento del estudiante, tiene riesgo de abandono de los cursos y se puede detectar las medidas para retener estudiantes.
- Visualización de datos: Los informes sobre datos educativos se vuelven cada vez más complejo a medida que los datos educativos crecen en tamaño. Las técnicas de visualización de datos para facilitar, identificar las tendencias y relaciones en los informes generados por esta técnica.
- Retroalimentación inteligente: Los sistemas de aprendizaje pueden proporcionar información inteligente e inmediata. Retroalimentación a los estudiantes en respuesta a sus aportes que mejorará la interacción y el rendimiento del estudiante.
- Recomendación del curso: Se pueden recomendar nuevos cursos a los estudiantes en función de los intereses de los estudiantes identificados mediante el análisis de sus ocupaciones. Eso asegurará que los estudiantes no este

la elección de su interés.

- Estimación de habilidades del estudiante: Habilidades adquiridas por el estudiante.

Detección de comportamiento: Comportamiento en la comunidad, actividades o juegos que ayudan a desarrollar a un estudiante un modelo de agrupación y colaboración de estudiantes, análisis de redes sociales, desarrollo de mapas conceptuales, construcción de material didáctico, planificación y programación.

Cortés (2016) explicó que hacer un rastreo sobre las implicaciones de los datos en las organizaciones es una oportunidad para entender qué está haciendo el sector privado y público en temas de *Big Data*, qué papel juega la Ciberseguridad en el futuro de la gestión de grandes cantidades de datos y, por supuesto, conocer algunos ejemplos de empresas y herramientas que son referencia a nivel mundial. Todas estas sensaciones son producto de su participación en Futurizz 2016.

## Capítulo 4. Metodología de la investigación

En este capítulo se describe la metodología que se va a utilizar en la elaboración de una propuesta para el cumplimiento de los objetivos planteados en este proyecto. El estudio se realizará en las oficinas del ICETEX, específicamente en la oficina de la Dirección de Tecnología, donde se encuentran las bases de datos que tienen la información para la ejecución del trabajo de investigación.

### 4.1 Tipo de investigación

El tipo de modalidad para este trabajo de grado es de carácter descriptivo con un enfoque mixto, ósea la combinación de los enfoques cualitativo y cuantitativo; en la primera parte se realizó una encuesta a los expertos de la dirección de tecnología y se utilizaron las bases de datos que reposan en el ICETEX.

Danhke citado por Batista y otros autores (2003, p.117), señala que los estudios descriptivos buscan especificar las propiedades, las características y los perfiles importantes de personas, grupos, comunidades o cualquier otro fenómeno que se someta a un análisis, de esta manera se toma este tipo de investigación. Y de igual manera Baptista, Fernández y Hernández (1991, p.21) señalan que los diseños mixtos representan el más alto grado de integración o combinación entre los enfoques cualitativo y cuantitativo.

El primer paso, ha sido evaluar diferentes tipos de metodologías de alineamiento estratégico de Big Data, a saber, el estudio de las 4V (Volumen, velocidad, variedad y valor) que realiza el modelo de Oracle, pero para esta investigación SAVI BIGD va más allá del análisis de las 4V. Este modelo integra el negocio con las tecnologías y es la más adecuada para las características del ICETEX.

SAVI-BIGD, se basa en un enfoque estratégico identificando el valor que tiene *Big Data* en alguna organización. Claro está, que se debe evaluar el problema mediante un estudio cualitativo inicialmente con las partes interesadas. Por lo cual el problema puede ser analizado a través de métodos de recolección de información como son las entrevistas y las encuestas. Para una primera fase se realizará encuestas a los directivos y personal del área de tecnología ya que ellos tienen una visión general y pueden tener un análisis de los datos que cuenta la entidad. A continuación, se describe literatura de cómo abordar estas encuestas y el enfoque de los resultados que se pueden entregar.

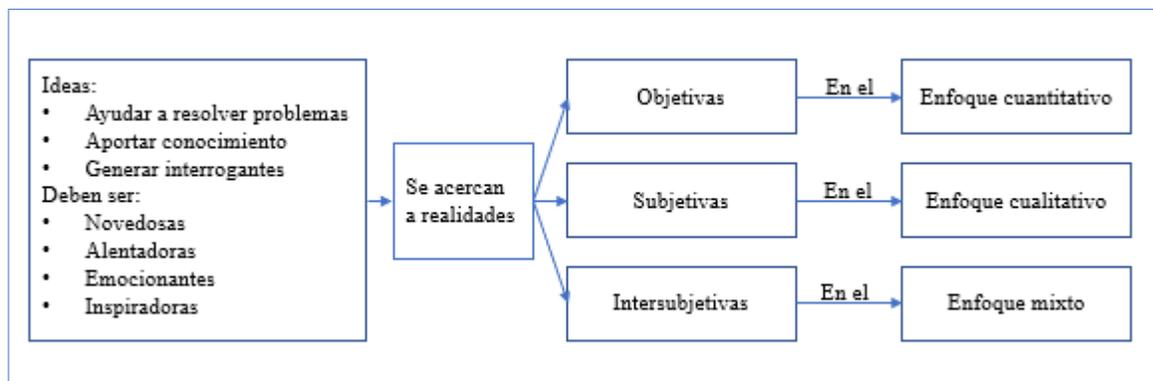
Ragin Charles C., en el capítulo 6 del libro *Metodologías de la investigación*, titulado “La construcción de la investigación social, Introducción a los métodos y su diversidad”, explicó cómo los métodos cuantitativos se concentran directamente en las relaciones entre variables

(dependientes o independientes), para mostrar un resultado. Este método, en comparación con los de corte cualitativo, sacrifica el conocimiento en profundidad. No obstante, para la presente investigación se tendrá en cuenta la explicación sobre los métodos cualitativos, ya que estos métodos identifican los diferentes patrones del comportamiento de la humanidad que impactan el pensamiento según los acontecimientos en las diferentes ciudades y países.

El uso de métodos cuantitativos permite que los investigadores generen ideas en cada etapa y puedan recoger información que le pueda servir para comprobar teorías, si se encontraron correlaciones entre las variables, las pruebas y las investigaciones siguen o creen patrones.

Según Charles, para realizar cualquier estudio, tipo o individuo, debe existir una guía creada en el pasado, pues estas servirán para organizar marcos analíticos para las diferentes áreas que generan un conocimiento para la creación de una estrategia. Por su parte, Danhke (1989), afirmó que los métodos cualitativos son de carácter descriptivo, pues buscan especificar las características, propiedades y perfiles importantes de personas, grupos, comunidades o cualquier otro fenómeno que se puedan analizar. Los estudios descriptivos permiten recoger información de manera independiente o en conjunto sobre las variables que se escojan.

Figura 12. Enfoque de proyectos



Fuente: Elaboración propia a partir de: Hernández, Fernández, y Baptista (1991).

En la figura 12, se observa características que acercan la realidad al enfoque metodológico que se puede escoger en un proyecto. Por ello, las variables cualitativas y cuantitativas aportan información a esta investigación que se complementan y una vez realizando y entendiendo estos métodos. La metodología que se quiere implementar en la captura de datos permite la recolección, el análisis e integración de las variables, para que estas generen resultados a este proyecto es de carácter descriptivo con enfoque mixto.

Por lo anterior y con los datos que tiene el ICETEX, este trabajo se realizará en un enfoque mixto, ya que se pueden revisar factores cualitativos y cuantitativos que ayudan a generar ideas para resolver problemas, aportar conocimiento y generar interrogantes de una forma novedosa que inspire a las diferentes áreas alinear sus objetivos estratégicos basados en las herramientas que suministra el uso del *Big Data*.

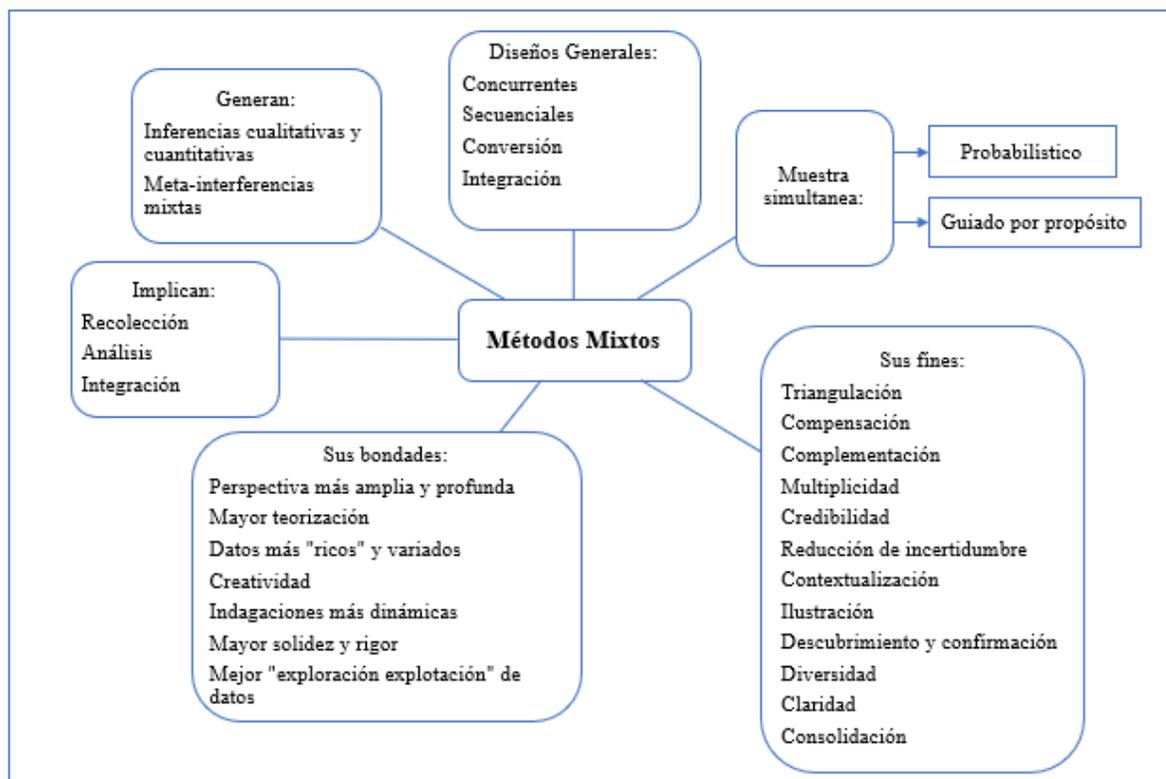
## 4.2. Técnica de recolección de datos

Las variables cualitativas y cuantitativas que se presentan en la figura 10 dan cuenta del conjunto de procesos que implican esta investigación.

Los métodos mixtos tienen un enfoque profundo en las diferentes áreas del ICETEX ya que cada una tiene diferentes necesidades para el cumplimiento de metas. Por ejemplo, el área comercial y mercadeo requiere capturar la satisfacción de todos los usuarios de alguno de los productos que han sido beneficiados, ya sea para realizar un cambio de imagen o fidelizar la satisfacción de los usuarios.

De igual manera como lo muestra la figura 13, las bondades de los métodos mixtos pueden generar una perspectiva más amplia sobre los datos analizados, mayor explotación y exploración de los datos que se tienen en el ICETEX.

Figura 13. Métodos mixtos



Fuente: Elaboración propia a partir de: Hernández, Fernández y Baptista (1991).

Hernández, Fernández y Baptista (1991, 546) señalaron que los métodos mixtos representan un conjunto de procesos sistemáticos, empíricos y críticos de una investigación e implican la

recolección y el análisis de datos cuantitativos y cualitativos, así como su integración y discusión conjunta, para realizar inferencias producto de toda la información recuperada.

### **4.3. Instrumentos y Fuentes**

El ICETEX es una entidad financiera que durante sus años de servicio ha logrado recolectar los datos de aproximadamente 744.000 beneficiarios, motivo que le ha llevado a generar una amplia fuente de información. Para esta investigación se utilizarán dos tipos de fuentes de información: Fuente participante: Encuesta realizada a los colaboradores de la Dirección de Tecnología del ICETEX; y la fuente no participante: Toda la información que se encuentra en las bases de datos del ICETEX que orientan esta investigación y tienen relación directa con la entidad y el negocio.

### **4.4. Instrumentos**

Entre los instrumentos seleccionados para este estudio se pretende brindar y captar información de los expertos que hacen parte de la Dirección de Tecnología mediante la encuesta, una técnica de recolección de datos por medio de un cuestionario. En esta función los cuestionarios cerrados son instrumentos valiosos en una investigación cuantitativa, que van a proporcionar datos estadísticos para complementar con los datos cualitativos.

El instrumento de encuesta, para esta investigación, tiene como finalidad recoger la percepción de los trabajadores sobre los datos con los que cuenta ICETEX y dar entendimiento a los datos numéricos que generan las bases de datos; pues gracias a estos se podrá analizar y generar propuestos de alto impacto en la entidad.

#### 4.4.1 Encuesta

Según García (2013), una encuesta es una técnica que utiliza un conjunto de procedimientos de una investigación, mediante los cuales se recoge y analiza una serie de datos sobre una muestra de sujetos representativa de una población. Gracias a las respuestas de la población que participa, se pretende explicar una serie de características.

Una de ellas puede ser los cuestionarios de interrogación, con el fin de obtener mediciones cuantitativas de una gran variedad de características objetivas y subjetivas de la población. Utilizando este instrumento se puede establecer un diseño de investigación, una hipótesis clara, definición de variables, diseño de cuestionario, organización de trabajo de campo, obtención y tratamiento de datos, análisis de datos e interpretación de resultados.

Por lo anterior se realiza una encuesta a los funcionarios y contratistas de la dirección de tecnología del ICETEX. Esta consta de quince preguntas cerradas en cinco segmentos de tres preguntas.

#### 4.4.2 Cuestionario

Para dar continuidad al instrumento de la encuesta, se debe elaborar el cuestionario. Este puede definirse como el documento de preguntas que tiene el contexto de la investigación a realizar. De igual manera hay que tener en cuenta las características de la población a la que se va a encuestar, número de preguntas, entre otras.

Según Gasperi (2003, 6), los cuestionarios deben ser operativos, fidedignos y válidos; por ello, deben ser sencillos, con preguntas que tengan variedad de opciones de respuesta y que sean cortos y entendibles. En este sentido, para este cuestionario, se definió preguntas dicotómicas, preguntas cerradas con respuestas prefijadas SÍ/NO.

### 4.5 Identificación de población

Puesto que una de las muestras de este estudio es el personal de la Dirección de Tecnología, primeros afectados por falta de una buena implementación de los datos, es necesario establecer un límite de personas a encuestar.

La otra Fuente: de estudio son los datos que reposan en las bases de datos del ICETEX, como los formularios de solicitud y los usuarios con cartera activa. Estos constan de aproximadamente 744.000 usuarios activos.

#### 4.5.1 Muestra bases de datos (BD) ICETEX

Para esta muestra se tiene en cuenta los datos que reposan en las BD del ICETEX que vienen siendo las fuentes de datos que se tendrán en cuenta en una de las fases del marco estratégico SAVI-BIGD y los diferentes informes que se generan para las diferentes áreas como la cantidad de usuarios inscritos para las diferentes líneas de crédito-Vicepresidencia de Crédito, líneas de fondos- Vicepresidencia de Fondos, Becas y convenios Internacionales - Oficina de Relaciones Internacionales, usuarios con cartera activa. Los usuarios de cartera activa se encuentran relacionados con las áreas de Dirección de Cobranza, Vicepresidencia de Operaciones, Vicepresidencia Financiera, desertores que tocan por lo menos tres de estas áreas.

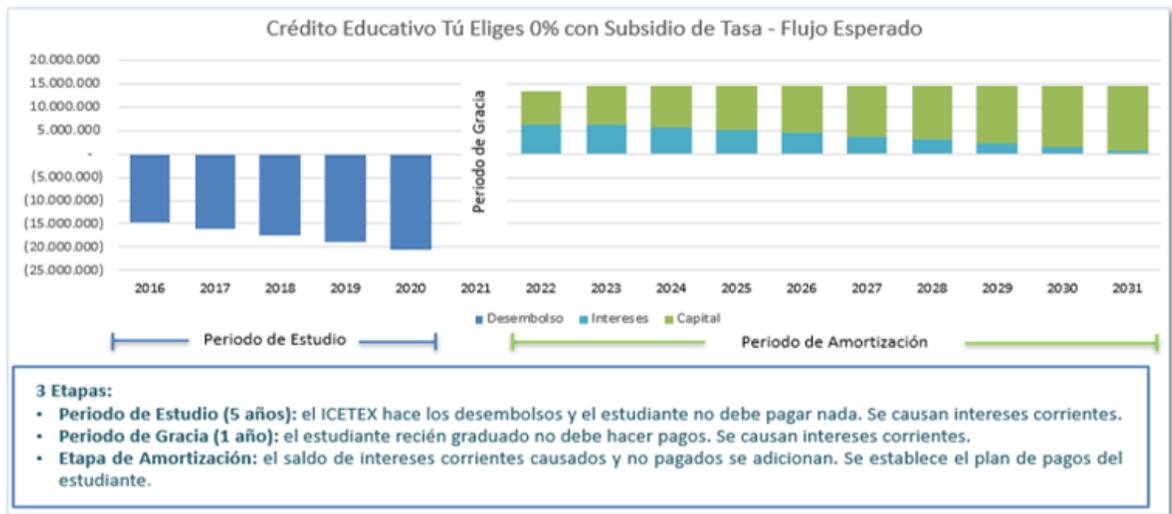
Ahora bien, la Oficina de Comercial requiere informes a diario para realizar diferentes campañas de fidelización en conjunto con la Oficina de Comunicaciones. A continuación, se realiza un resumen de los diferentes informes que se generan para las áreas a partir de los datos que se encuentran en ICETEX y hacen parte de la investigación.

El ICETEX otorga anualmente diferentes productos a través de créditos. Después del otorgamiento se pasa al cobro con la asignación de un plan de pagos, con una estructura única, para cubrir y acompañar a los beneficiarios financieramente desde el inicio hasta el fin

de los estudios. Una vez se cursa la totalidad del programa académico y los estudiantes se gradúan, inician los pagos para darle la oportunidad a otro estudiante.

Dada esta información se ven involucradas las diferentes áreas del ICETEX, pero en una fase inicial que toca uno de los objetivos estratégicos que es “adjudicación de créditos educativos”. Siendo un macroproceso de la vicepresidencia de crédito y cobranza y este a su vez alimenta con insumos a otras áreas en diferentes casuísticas, como son cartera vigente, para graduados y desertores. En la figura 14, se puede apreciar la estructura de un crédito educativo que detalla los beneficios que ofrece el ICETEX, un tipo de crédito con beneficios únicos que se amoldarían a evitar deserción por lo menos en los casos que no tienen dinero para iniciar, mantener y graduar.

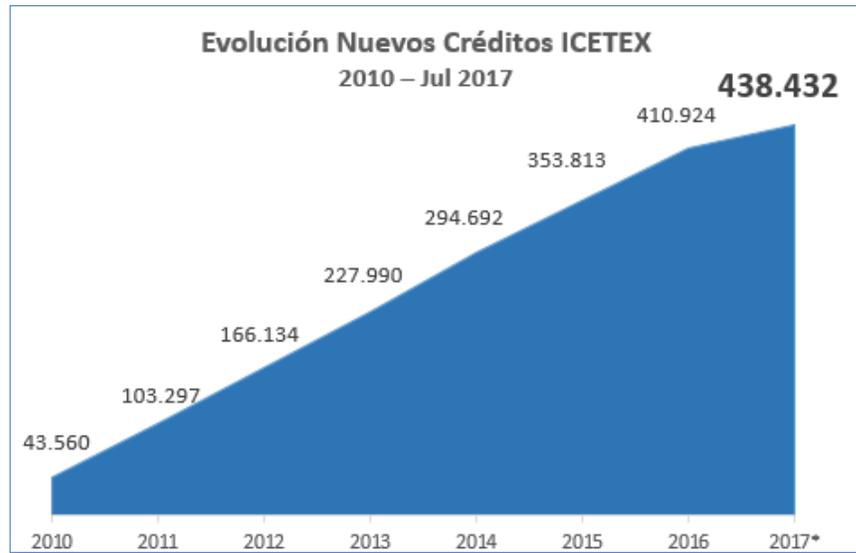
Figura 14. Estructura crédito ICETEX



Fuente: Oficina Asesora de Planeación ICETEX 2017

En el informe de 2018 del ICETEX, en la figura 14, se puede observar la sumatoria por año de la evolución de los usuarios inscritos para las líneas de crédito del ICETEX.

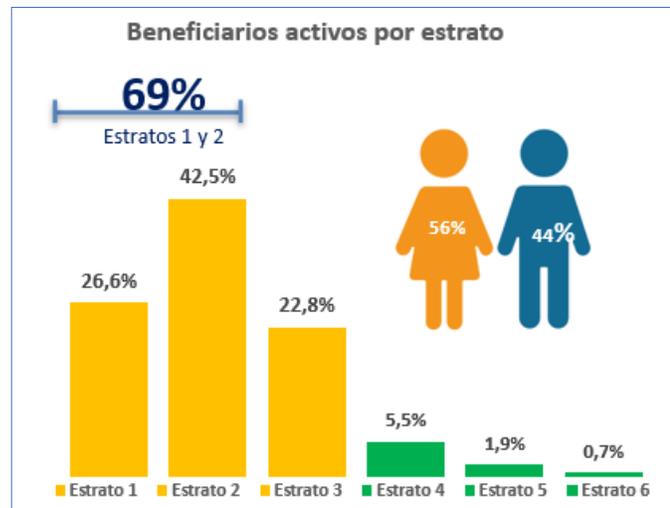
Figura 15. Evolución créditos nuevos



Fuente: Informe 2018 ICETEX

Como se ha mencionado, analizar estos datos e identificar qué características pueden presentar estos usuarios, de qué estratos pertenecen y cómo contribuyen a la educación superior y de ahí a diferentes sectores una vez logran culminar sus estudios, de igual manera el porcentaje por género es más alto en las mujeres que para los hombres y el estrato 2 tiene mayor porcentaje de los usuarios con crédito educativo, en la figura 15, se puede observar el porcentaje de usuarios activos según su estrato socioeconómico y género.

Figura 16. Usuarios activos por estrato y género



Fuente: Informe 2018 ICETEX

Por otro lado, este informe presenta la deserción académica por semestre. El Banco Mundial en el año 2018 publicó el análisis de estos datos e indicó que Colombia es el segundo país en América Latina con mayor tasa de deserción universitaria. Se estima que el 42 % de los que ingresan a planteles educativos terminan desertando en los primeros años, esto a noviembre de 2018. Sin embargo, revisando el informe del 2016, se evidencia que en los primeros tres semestres de vida universitaria es más alto el porcentaje de deserción. En el reporte de SPADIES, en la tabla 13, se indican los siguientes datos:

*Tabla 13. Tabla deserción por semestre 2016*

Semestre	1	2	3	4	5	6	7	8	9	10	100%
Promedio	37%	17%	12%	8%	6%	5%	5%	4%	3%	3%	

Fuente: Elaboración propia construida a partir de datos del reporte SPADIES

Ahora, se puede observar en la tabla 14 un porcentaje de deserción por departamentos y por ciclos propedéuticos (Página Ministerio de Educación , 2009); estos últimos creados bajo la “Ley 749 de 2002 que introduce en el sistema educativo la formación por ciclos propedéuticos en las áreas de ingenierías, tecnología de la información y administración”.

Los ciclos propedéuticos estas son unidades interdependientes, complementarias y secuenciales que facilitan a estudiantes el ingreso a la educación superior y así continuar en el proceso de formación a lo largo de la vida, en este caso particular, en el pregrado.

En consecuencia, un ciclo propedéutico se puede definir como una fase de la educación que le permite al estudiante desarrollarse en su formación profesional siguiendo sus intereses y capacidades. Ahora, si estos ciclos facilitan el ingreso a estudios superiores, también se observa que la deserción tiene un porcentaje aproximado del 18% por ciclos, como se puede revisar en la tabla 14. Datos que presentan diferentes características por ciudades y por nivel académico.

*Tabla 14. Porcentaje de deserción por departamento y ciclo propedéutico*

Tasa de Deserción Anual por Nivel de Formación y Departamento Año 2016					
Departamento	Técnica Profesional	Tecnológica	TyT Agregado	Universitaria	Total
Amazonas	n.d.	n.d.	n.d.	n.d.	Sin registro
Antioquia	30,59%	17,10%	15,52%	9,69%	18,23%
Arauca	n.d.	n.d.	n.d.	n.d.	Sin registro
Atlántico	43,20%	20,54%	26,74%	9,11%	24,90%
Bogotá	25,99%	19,78%	23,04%	9,37%	19,55%
Bolívar	28,51%	15,61%	21,89%	9,46%	18,87%

Boyacá	n.d.	27,27%	27,05%	7,05%	20,46%
Caldas	25,72%	19,64%	13,52%	5,74%	16,16%
Caquetá	n.d.	30,43%	30,43%	7,51%	Sin registro
Casanare	n.d.	n.d.	n.d.	14,57%	14,57%
Cauca	8,11%	16,27%	14,51%	7,32%	11,55%
Cesar	47,37%	15,28%	47,37%	11,73%	30,44%
Chocó	n.d.	28,57%	28,57%	11,79%	22,98%
Córdoba	n.d.	7,93%	7,93%	6,76%	7,54%
Cundinamarca	35,71%	5,42%	5,13%	6,61%	13,22%
Guainía	n.d.	n.d.	n.d.	n.d.	Sin registro
Guaviare	n.d.	n.d.	n.d.	n.d.	Sin registro
Huila	29,85%	15,09%	15,52%	8,63%	17,27%
La Guajira	35,52%	7,87%	23,52%	9,87%	19,20%
Magdalena	29,03%	20,97%	25,29%	8,60%	20,97%
Meta	n.d.	14,47%	14,47%	8,35%	12,43%
Nariño	15,79%	9,55%	10,20%	7,02%	10,64%
Norte de Santander	34,54%	16,89%	12,78%	8,97%	18,30%
Putumayo	n.d.	20,48%	20,67%	14,98%	18,71%
Quindío	23,88%	20,82%	22,27%	8,52%	18,87%
Risaralda	16,27%	15,39%	15,67%	8,05%	13,85%
San Andrés, Providencia y Santa Catalina	n.d.	n.d.	56,25%	n.d.	56,25%
Santander	31,59%	22,34%	18,26%	9,01%	20,30%
Sucre	9,76%	10,38%	11,94%	9,18%	10,32%
Tolima	23,13%	19,21%	20,55%	9,29%	18,05%
Valle del Cauca	18,25%	17,06%	17,13%	9,31%	15,44%
Vaupés	n.d.	n.d.	n.d.	n.d.	Sin registro
Vichada	n.d.	n.d.	n.d.	n.d.	Sin registro
<b>Total, Nacional</b>	<b>26,13%</b>	<b>16,70%</b>	<b>17,10%</b>	<b>9,03%</b>	<b>18,81%</b>
Fuente: SPADIES Fecha de corte marzo de 2017					

En los datos de las tablas anteriores, se puede observar diferentes variables como son departamentos, estratos socioeconómicos. Estos, a su vez, están relacionados con núcleos familiares, edades de los jóvenes que desean estudiar, tipo de estudio profesional, entre otros. Estos datos generan ideas de conjuntos y subconjuntos que se pueden estudiar a fondo.

En la tabla 15, se observa el porcentaje de deserción de acuerdo con las áreas de conocimiento y con los ciclos propedéuticos, en donde el área de las ciencias de la educación presenta el porcentaje más alto:

Tabla 15. Deserción anual por área y nivel de formación

Área del conocimiento	Técnico Profesional	Tecnológica	Universitaria	Total
Ciencias de la educación	41,18%	12,14%	9,60%	63%
Economía, administración, contaduría y afines	28,36%	16,53%	9,64%	55%
Ingeniería, arquitectura, urbanismo y afines	26,67%	17,86%	9,25%	54%
Ciencias de la salud	26,02%	13,51%	6,66%	46%
Bellas Artes	18,02%	19,31%	8,26%	46%
Ciencias sociales y humanas	20,35%	14,68%	10,41%	45%
Agronomía, veterinaria y afines	17,71%	16,41%	9,19%	43%
Matemáticas y ciencias naturales	17,33%	12,12%	11,37%	41%

Fuente: SPADIES, fecha corte marzo 2017

En la información anterior, existe población de las diferentes áreas académicas, de los diferentes estratos socioeconómicos y otras características. Como se evidencia en la encuesta a los funcionarios, se puede analizar la falta de agilidad para alinear esta información a lo que se quiere en el ICETEX, como la asignación de cobertura a nivel nacional, la implementación de herramienta en *Big Data*, analítica y demás tecnologías modernas que actualmente no están.

#### 4.6. Descripción de la muestra

La muestra se aplica a los posibles interesados como son los ingenieros vinculados laboralmente a la dirección de tecnología, pues ellos se podrán ver beneficiados con la creación de un grupo especialista en datos y que ayuden a reducir los tiempos en las actividades realizadas en esta área.

#### 4.7. Diseño de los instrumentos

Se elaboro un cuestionario de 15 preguntas cerradas en 5 segmentos, cada uno de ellos con tres preguntas elaboradas bajo el marco de arquitectura basada en *Big Data* SAVI-BIGD, que tiene cinco fases (visión estratégica, hoja de ruta de implementación para la fase BIGD, generación de objetivos estratégicos Big BIGD, determinación de fuentes de datos e implementación de *Big Data*). Se pretende analizar la percepción de los funcionarios de la Dirección de Tecnología sobre la importancia de ordenar los datos y cómo podrían ser involucrados en el plan estratégico de la entidad, en la planeación de la infraestructura y en las fuentes de datos como son:

- Las fuentes endógenas: aquellas que se revisan desde el interior de la entidad.

- Las fuentes exógenas: aquellas para realizar estrategias hacia el exterior de la entidad involucrando las áreas de comercial y comunicaciones.

#### 4.8. Aplicación de instrumentos

El cuestionario se realizó en la herramienta de Google Forms, que sirvió como insumo para la captura de los datos de cuestionario. La duración de esta es de aproximadamente cinco-diez minutos. Estas se realizaron en el mes de abril de 2020.

#### 4.9. Tabulación

En la tabla 16 se realiza tabulación sobre los resultados obtenidos en las encuestas realizadas.

*Tabla 16. Tabulación resultados encuesta*

Segmento	Dimensión	Encuesta	Cuestionario	Propósito
1	Visión Estratégica, contexto y alcance del negocio.	43	3	Identificar la percepción del conocimiento del negocio y el uso de la tecnología con los datos históricos del ICETEX
2	Hacia dónde dirigir la implementación del buen uso de los datos	43	3	Identificar la percepción de los funcionarios y su posible aporte a partir de su conocimiento y estudio.
3	Estado actual de los datos y posibles usos	45	3	Analizar el sentido de pertenencia del trabajador del área de tecnología y el conocimiento de los aplicativos del ICETEX.
4	Aplicación y visión futura del procesamiento de datos	45	3	Analizar la capacidad y la intención colaborativa con los procedimientos que tiene la entidad para innovar y permitir la participación de los funcionarios.
5	Cómo monetizar las BD y qué barreras para lograr alinearlas al plan estratégico del ICETEX	45	3	Identificar el conocimiento de las herramientas que tiene el mercado para la implementación de analítica de datos o implementación de <i>Big Data</i> .
<b>Total, preguntas</b>		<b>221</b>	<b>15</b>	

Fuente: Elaboración propia, resultados encuesta

## Capítulo 5. Análisis de instrumento y propuesta

En este capítulo, se pretende analizar las diferentes variables que se aplicaron en las encuestas o instrumento seleccionado y la revisión de diferentes poblaciones de las bases de datos. Estos resultados se presentarán por medio de ilustraciones y tablas para lograr interpretar y analizar. El diseño de las preguntas se tomó en cinco segmentos que resumen tres modelos de arquitectura basada en análisis de datos. En la tabla 17, se mencionan los modelos.

*Tabla 17. Arquitecturas basadas en datos*

Arquitecturas basadas en Datos		
Oracle Enterprise Architecture Development Process	Data-Driven Business Models	El marco SAVI-BIGD

Fuente: Elaboración propia

Estas arquitecturas definen pasos importantes para tener en cuenta en un alineamiento estratégico basado en datos. Los resultados obtenidos pretenden analizar la percepción de los participantes según los datos existentes en el ICETEX y cómo se podrían utilizar para el mejoramiento de procesos o posibles estrategias que incentive el sector de la educación.

En este caso, se utilizó el instrumento de la encuesta que tienen relación directa con el trabajo de la dirección de tecnología del ICETEX. Estas encuestas se enviaron a los 58 integrantes de esta dirección; sin embargo, solo realizaron encuestas a 43 funcionarios y contratistas que se encuentran ubicados en las diferentes coordinaciones y células de trabajo de esta dirección. En la tabla 18, se presenta los perfiles de los encuestados.

*Tabla 18. Integrantes dirección de tecnología*

Dirección de Tecnología		
Coordinación Sistemas de Información	Coordinación de Operaciones	Coordinación de Infraestructura
Coordinador	Coordinador	Coordinador
Analistas 5	Ingenieros 10	Mesa de Servicio 10
Desarrolladores 10	Soporte 5	Ingenieros Infraestructura 10
Ingenieros QA 5		

Fuente: Elaboración propia

### 5.1. Aplicación de la encuesta

La encuesta se aplicó en cinco segmentos, basada en el entendimiento del modelo de arquitectura basada en datos. 43 funcionarios de la dirección de tecnología realizaron la encuesta (Anexo 1), mediante la herramienta de Google Forms. Este método se organizó en

quince preguntas de cinco segmentos. El tiempo de diligencia fue de aproximadamente 15 minutos.

## 5.2. Propuesta

Una vez analizados los resultados obtenidos en las encuestas y en los diferentes informes que proporcionan las bases de datos del ICETEX, se evidencia la necesidad de alinear las tecnologías de información con los objetivos estratégicos del ICETEX. Por ello, se investigó tres marcos de arquitectura alineados a *Big Data* y el marco de referencia del Ministerio de Tecnologías de la Información y las Comunicaciones – MinTIC.

Como se mencionó en el capítulo 3, Lakoju (2017) en su trabajo *SAVI-BIGD (Strategic Approach Of Value Identification Big Data Framework)* crea este marco que adopta la construcción del diseño y evalúa la naturaleza del diseño y menciona que *Big Data* tiene el potencial de enormes beneficios para una organización; sin embargo, se requiere una inversión, tiempo e infraestructura. Se centran más en "cómo implementar *Big Data*" y "tipo de infraestructuras de *Big Data*", lo que significa para el ICETEX la integración de estas dos coordinaciones y la creación de un grupo que pueda alinear desde la estrategia de TI a una visión de toda la entidad.

## 5.3. Estrategia para propuesta marco alineado en *Big Data*

Las tendencias de innovación obligan a las instituciones a buscar alternativas, como la *Big Data*. Para ello, se pretende tomar el marco estratégico SAVI-BIGD, ya que combina teoría de alineación con la teoría estratégica comercial y digital y el resultado de estos dos genera el marco estratégico de *Big Data*, proyectos que ayudaron a las compañías a identificar el valor potencial a través de la implementación de *Big Data*. Por lo anterior, este marco le permitirá al ICETEX combinar la alineación estratégica, con las estrategias planteadas en lo comercial y digital permitiendo ser uno de los pilares en estas innovaciones tecnológicas. Por otro lado, se tendrá en cuenta las mejores prácticas que describe la guía técnica de información de gobierno de datos vigente del Ministerio de Tecnologías de la Información y las Comunicaciones – MinTIC.

- SAVI-BIGD Estrategia Framework (Lakoju and Serrano, 2016)
- Oracle Enterprise Architecture Development Process
- Modelos de Negocios basados en datos “Estudio Universidad Cambridge”(Data-Driven Business Models)

Estos tres modelos se caracterizan por el alineamiento estratégico basado en datos. Todos ellos tienen fases comunes. Una vez realizada la investigación de las arquitecturas alineadas a los datos se propone utilizar el Marco Estratégico SAVI-BIGD que consta de cinco fases:

### 5.4. Fase 1: Visión estratégica

Para esta fase se tendrá en cuenta la visión del ICETEX donde indica:

Seremos a 2025 la Entidad de los colombianos que contribuye significativamente a la transformación social y a una Colombia mejor educada, convirtiéndonos en uno de los 5 líderes internacionales en gestión de recursos de fomento a la educación, con eficiencia organizacional, orientación al cliente, innovación y apoyo tecnológico (ICETEX, 2016, 6).

Según la visión de ICETEX y aplicándolo a esta primera fase de la Visión estratégica se realizan las siguientes actividades:

Figura 17. Fase visión estratégica



Fuente: Elaboración propia

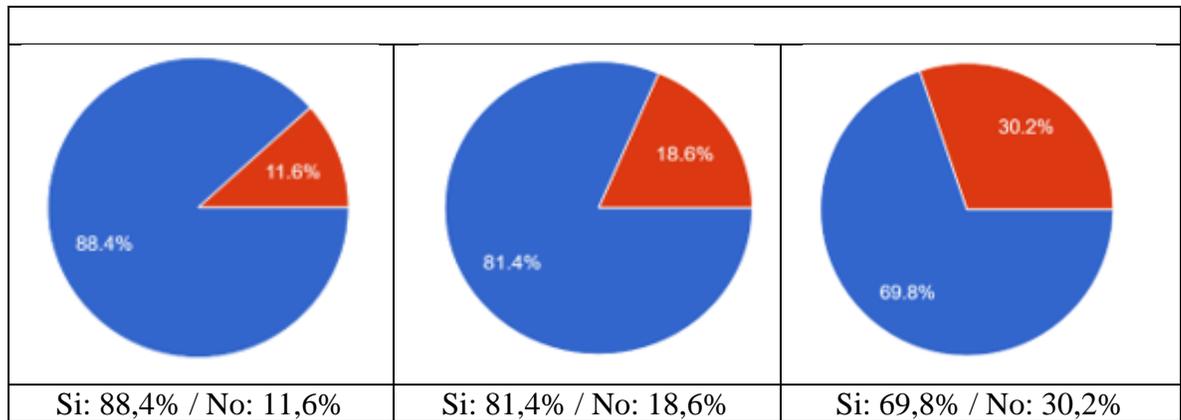
5.4.1. Aplicación y resultados de la encuesta de la siguiente manera:

- **Segmento 1 – Visión estratégica, contexto y alcance del negocio**

<b>Segmento 1 – Visión estratégica, Contexto y alcance del negocio.</b>
1. ¿Cree usted que el uso de los datos que reposan en el ICETEX puede ser utilizado en el plan estratégico?
2. ¿Las IES o demás constituyentes que tienen convenio con el ICETEX podrían aportar información que contribuya a la innovación o cambio de estrategia institucional?
3. ¿Considera usted que la información de las BD del ICETEX es consistente en todos los aplicativos?

Este segmento está diseñado en tres preguntas para analizar la percepción de los funcionarios con respecto a las BD en un alineamiento estratégico. En este, se observa un 79,8% que las BD del ICETEX podrían ayudar a contribuir a una mejora estratégica.

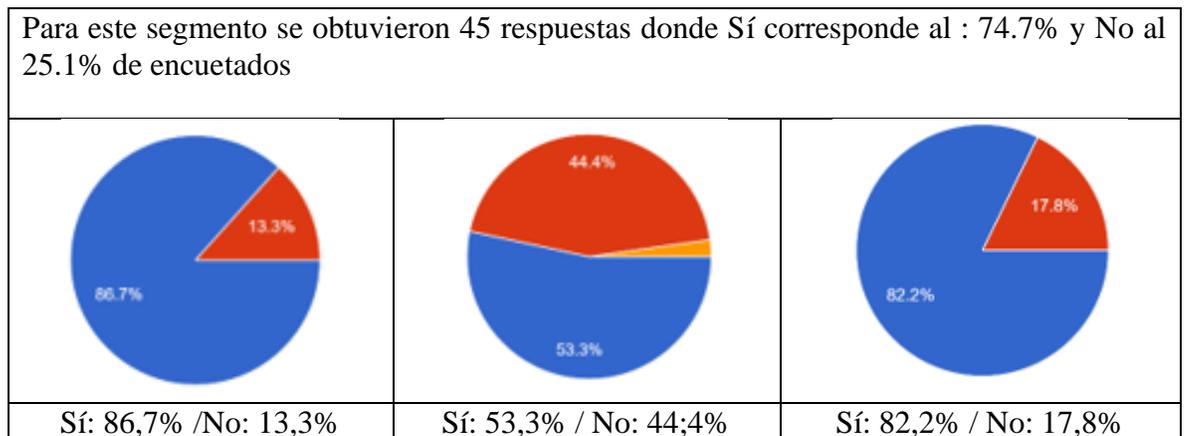
Para este segmento se obtuvieron 43 respuestas donde Si corresponde al: 79,8% y No al 20,2% de encuetados



- **Segmento 2 – Hacia dónde dirigir la implementación del buen uso de los datos**

<b>Segmento 2 – Hacia dónde dirigir la implementación del buen uso de los datos</b>
4. ¿Cree usted que los datos del ICETEX se pueden procesar para proponer una innovación tecnológica?
5. ¿Usted cree que la información de las bases de datos de los funcionarios del ICETEX (profesión, preparación o experiencia) contribuyen a las labores que ejercen en sus puestos de trabajo?
6. ¿Cree usted que se podría reubicar las áreas del ICETEX según las profesiones y experiencia laboral del universo de los funcionarios?

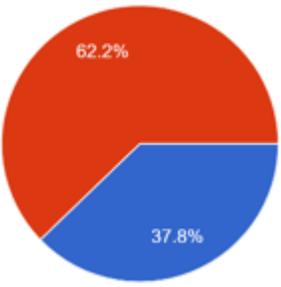
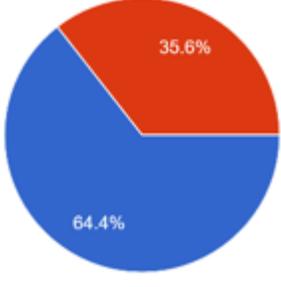
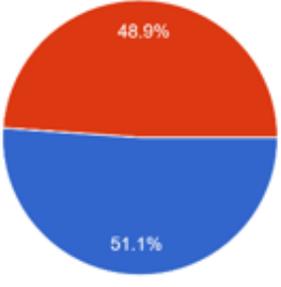
En este segmento, el 74,7% de los encuestados afirman que el buen uso de los datos que reposan en el ICETEX podría dirigir a la entidad en una innovación tecnológica y de igual manera organizar internamente a los funcionarios que tienen conocimiento en diferentes áreas.



**Segmento 3 – Estado Actual de los datos y posibles usos**

<b>Segmento 3 – Estado Actual de los datos y posibles usos</b>
7. ¿Usted cree que la información de las bases de datos del ICETEX podría contribuir en crear patrones sobre las posibles profesiones de los usuarios y cómo aportar a la empleabilidad el país?
8. ¿Usted, como ingeniero, podría evaluar el estado actual de las BD del ICETEX?
9. ¿Si usted tuviera una iniciativa para utilizar las BD del ICETEX en una posible modernización, cree usted que podría acceder a esa información?

Para este segmento, el 59,2% de los encuestados identifican el valor potencial que tienen las BD del ICETEX y opinan que ellos, desde su rol, pueden contribuir a un buen uso de datos y a la conformación del posible grupo que intervenga en una estrategia de modernización basada en análisis de datos.

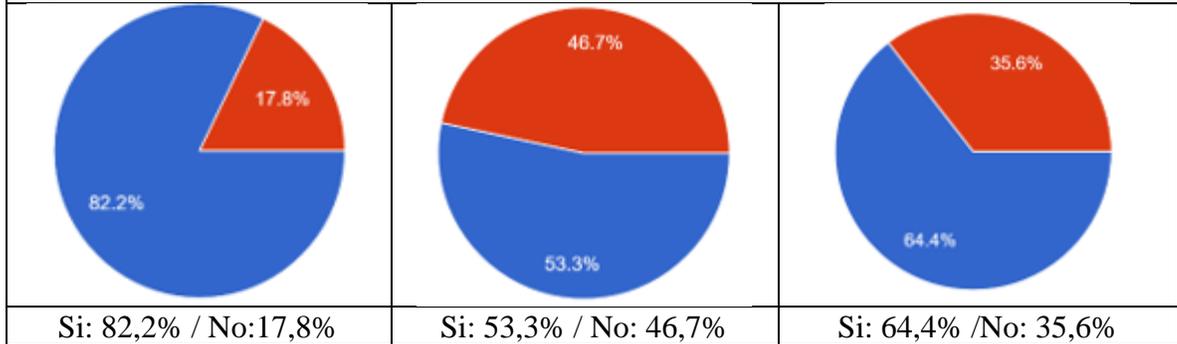
Para este segmento se obtuvieron 45 respuestas donde Sí corresponde al : 59,2% y No al 40,7% de encuestados		
		
Sí: 62,2% / No:37,8%	Sí: 64,4% / No: 35,6%	Sí: 51,1% /No: 48,9%

- **Segmento 4 – Aplicación y visión futura del procesamiento de datos**

<b>Segmento 4 – Aplicación y visión futura del procesamiento de datos</b>
10. ¿Cree usted que las diferentes entidades del estado y el ICETEX podrían integrar sus bases de datos y contribuir a la modernización del país?
11. De acuerdo con los procedimientos que existen en el ICETEX, ¿cree usted que son un impedimento para el trabajo diario?
12. ¿Con el personal activo, incluido contratistas, podría usted formar un grupo que gestione y organice la aplicación de las bases de datos del ICETEX?

Para este segmento, el 66,6% de los encuestados creen que el personal y el almacén de datos podrían contribuir a gestionar y organizar estrategias que se unan con demás entidades para plantear ideas a nivel nacional.

Para este segmento se obtuvieron 45 respuestas donde Si corresponde al : 66,6% y No al 33,3% de encuetados



- **Segmento 5- Cómo monetizar las BD y qué barreras hay para lograr alinearlas al plan estratégico del ICETEX**

#### Segmento 5- Cómo monetizar las BD y qué barreras para lograr alinearlas al plan estratégico del ICETEX

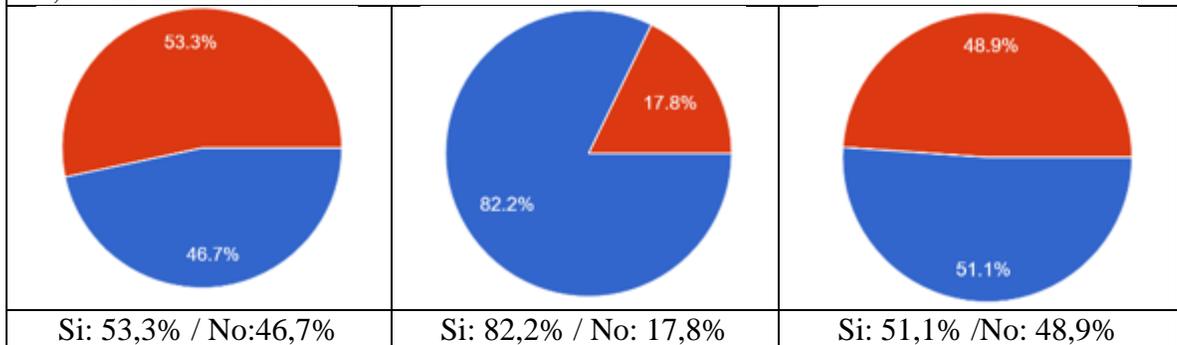
13. ¿Conoce alguna herramienta que pueda analizar los datos históricos del ICETEX para describir o predecir una posible estrategia institucional?

14. ¿Cree usted que las BD del ICETEX son útiles o podrían ser utilizadas en otras entidades?

15. Dado que el ICETEX es una entidad de nivel nacional, ¿cree usted que todos los sectores están involucrados y ayudan al cumplimiento de los objetivos estratégicos?

Para este segmento, el 60% creen que los sectores involucrados con el ICETEX podrían ayudar a alinear o a cumplir los objetivos estratégicos. Por otra parte, el 40% no tiene claro una herramienta que pueda realizar análisis patrones o definan un camino para implementar una arquitectura guiada por datos.

Para este segmento se obtuvieron 45 respuestas donde Si corresponde al : 62,2% y No al 37,8% de encuetados



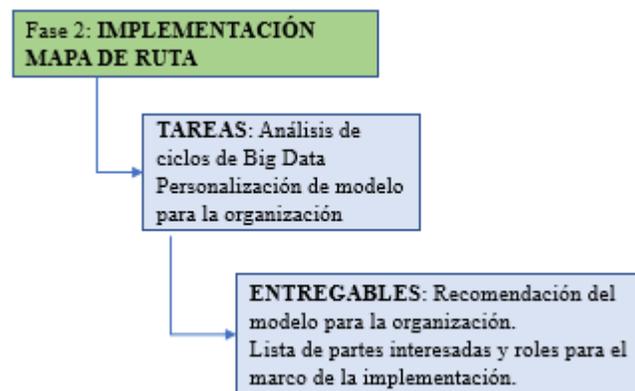
De igual manera en esta fase se tendrá en cuenta el objetivo de tecnología transversal en toda la entidad:

Elaboración de estrategia de trabajo (objetivo, herramienta, presentación y sensibilización). Entre los objetivos estratégicos del ICETEX se encuentra el objetivo A2: “convertir las tecnologías de información en una ventaja competitiva del negocio” (ICETEX, 2016, 18). Este objetivo es transversal en el mapa estratégico del ICETEX y por ello tiene en cuenta las mega tendencias tecnológicas como movilidad, Cloud, *Business Intelligence/analytics – Big Data* y *Social Media*.

### 5.5. Fase 2: Diseño de hoja de ruta para *Big Data*

Para esta fase se va a realizar el plan de trabajo e identificar áreas interesadas a las que se tendrán en cuenta inicialmente para esta estrategia y de igual manera los estándares propuestos por Ministerio de Tecnologías de la Información y las Comunicaciones – MinTIC, teniendo en cuenta los siguientes puntos:

Figura 18. Fase implementación mapa de ruta



Fuente: Elaboración propia

Priorización de áreas de trabajo (reportes, integración con coordinaciones de la VOT). Actualmente las vicepresidencias de crédito, vicepresidencia de fondos y la oficina de comercial y mercadeo y comunicaciones están atentos a los reportes para realizar campañas de fidelización y seguimiento a los estudiantes que tienen un producto con el ICETEX (beca, crédito, fondo). Se realizan aproximadamente 80 reportes mensuales, por lo cual se pretende suprimir estas solicitudes.

La identificación de herramientas de actualización de información hace que se recopile el estado actual del ICETEX en cuanto al manejo de los datos y se realizaran actualizaciones en las guías según los estándares y políticas de gobierno de datos, donde se tienen en cuenta el almacenamiento (*Storage*), seguridad (*Security*), calidad (*Quality*) e *interoperabilidad* (Intercambio de información) de los datos. De igual manera y con los servicios que tiene el ICETEX se apalancará el uso de las herramientas de Office 365 y sus históricos.

Para el análisis y necesidades de implementar la hoja de ruta de tiene contemplado el trabajo de los ingenieros del área de tecnología que actualmente están vinculados e identificados en la tabla 19. Equipo de trabajo:

*Tabla 19. Equipo de trabajo*

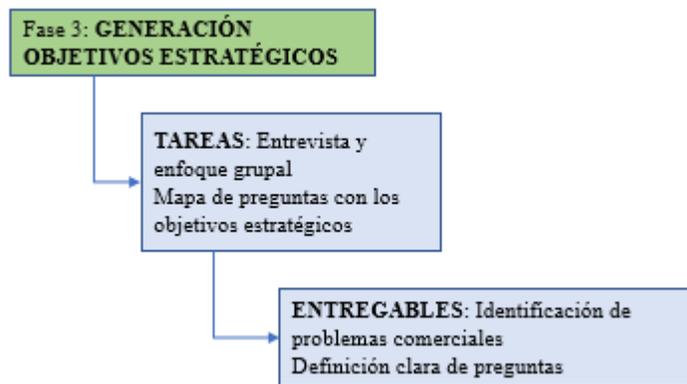
Jerarquía
Coordinador de Datos
Arquitecto de Datos
Científico de Datos
Ingeniero de Datos
Expertos en herramientas de <i>Big Data</i>
Analista de Datos
Ingeniero de Calidad

Fuente: Elaboración propia

### 5.6. Fase 3: Generación de Objetivos Estratégicos de *Big Data*

Para esta fase se deben tener en cuenta los objetivos que se van a trabajar y las problemáticas que se presentan en el área comercial.

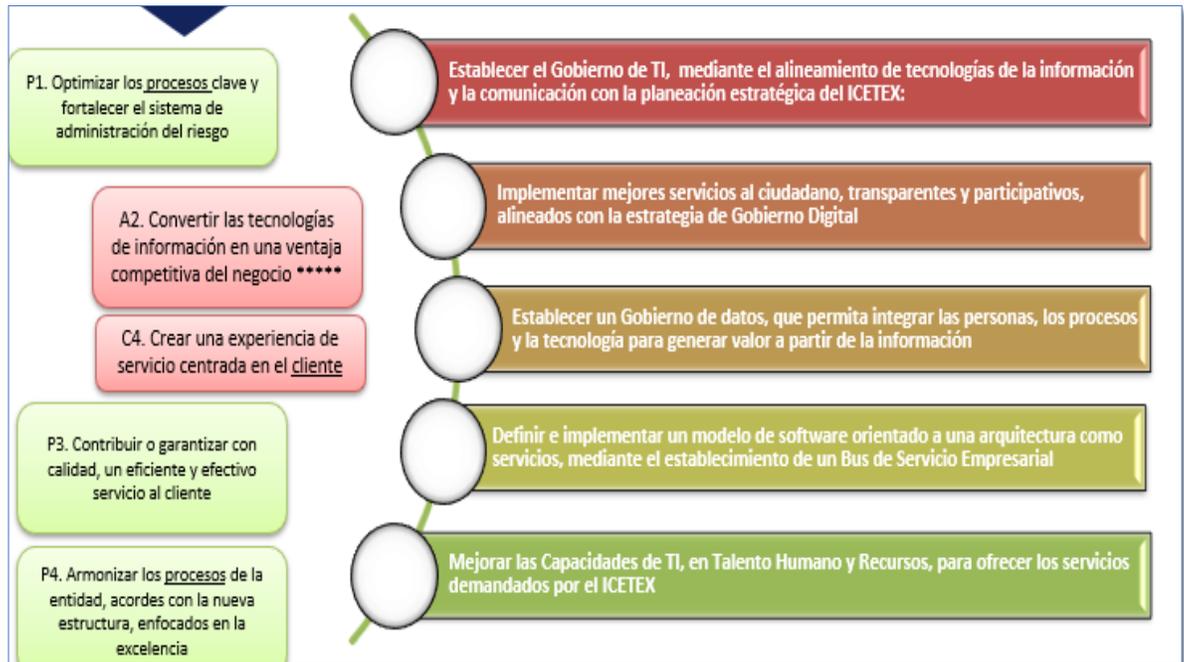
*Figura 19. Fase generación objetivos estratégicos*



Fuente: Elaboración propia

Como se mencionó en la fase 1, se tendrá en cuenta el objetivo: "convertir las tecnologías de información en una ventaja competitiva del negocio, este objetivo se encuentra transversal en el mapa de estratégico del ICETEX y pretende optimizar, contribuir y crear experiencias a los clientes del ICETEX (ICETEX, 2016, 18). Este está relacionado con las áreas en los siguientes objetivos:

Figura 20. Objetivos estratégicos de tecnología



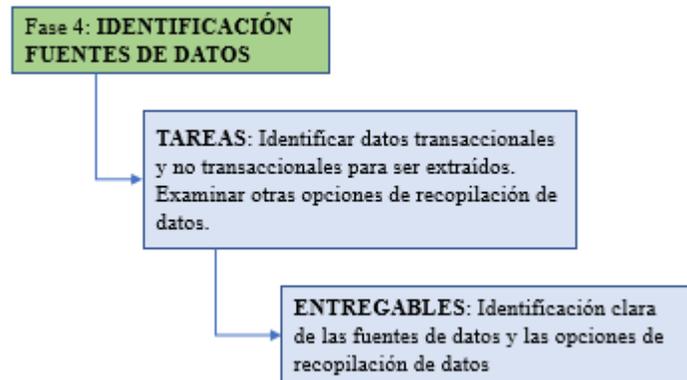
Fuente: Elaboración de la dirección de tecnología, 2019

Dentro del mapa estratégico de la Vicepresidencia de Operaciones y Tecnología, y de la mano con la Dirección de Tecnología, se pretende que esta área deje de ser un área de apoyo y se convierta en parte fundamental para definir una estrategia institucional.

#### 5.7. Fase 4: Identificación de fuentes de datos

En esta fase se tendrá en cuenta el estándar implementado por MinTIC en el ciclo de la vida del dato:

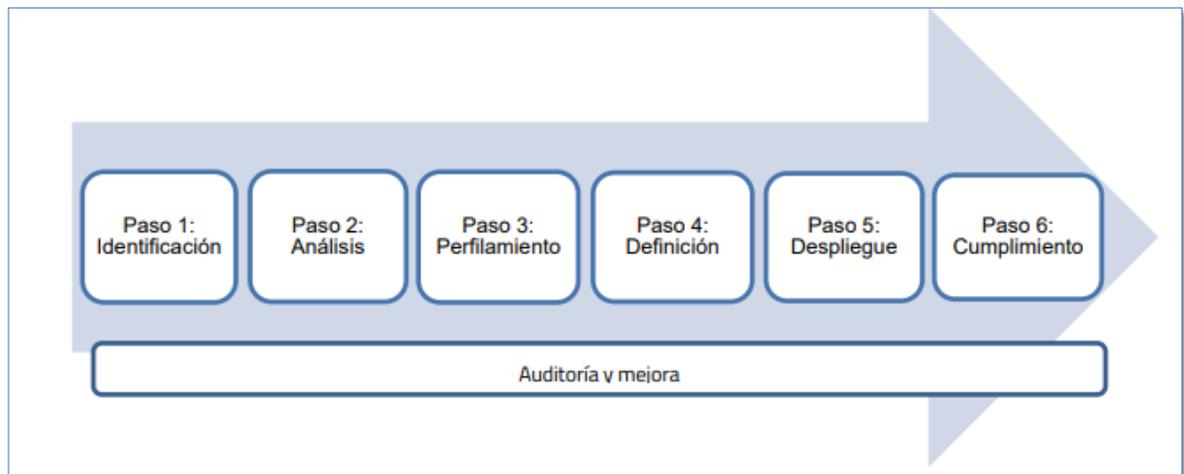
Figura 21. Fase identificación fuentes de datos



Fuente: Elaboración propia

La figura 22 muestra seis pasos que se deben tener en cuenta y que aporta a los entregables de esta fase en la identificación de las fuentes de los datos.

Figura 22. Pasos para desarrollar e implementar la gestión del ciclo de vida del dato



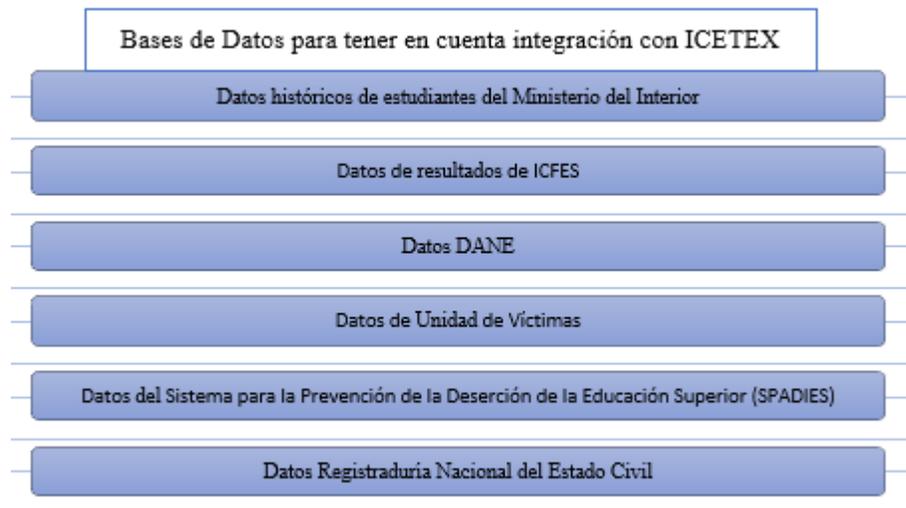
Fuente: Ministerio de Tecnologías e Información, 2014, 22.

En la comprobación de los seis pasos del ciclo de vida del dato y las dimensiones de la calidad de datos se requiere verificar que se haya actuado acorde al plan, identificando el uso de los datos en los directorios, entendimiento de los metadatos, datos maestros ubicación en los sistemas de información para reducir los niveles de inconsistencias y duplicidad en los datos, siendo esta una problemática identificada. De igual manera la revisión de que la información está disponible y sin datos ausentes para la exactitud de esta, datos correctos sin errores con valores uniformes, consistentes e integridad para tener una buena conectividad e integración con los sistemas de información y utilización de los recursos de infraestructura.

### 5.7.1. Fuentes de búsqueda y posible integración

Dentro de las fuentes de datos, estas son otras posibles opciones que aportan información para complementar dicha investigación.

*Figura 23. Fuentes de datos*



Fuente: Elaboración propia

De las bases de datos mencionadas, se realizará una integración para no tener duplicidad y unificar los registros de los usuarios, y así poder analizar y entender las necesidades de las familias beneficiadas por un crédito ICETEX. Los usuarios de ICETEX que ya reportan en las bases del Ministerio una profesión financiada y su posible empleabilidad una vez se hayan graduado, las posibles causas de la deserción y finalmente el comportamiento de la cartera del ICETEX, entregar un criterio para la toma de decisiones.

Una de las bases de datos que se debe revisar es la del DANE, especialmente, las personas que acceden a la educación superior se gradúan y no logran emplearse; esto puede generar un patrón para una estrategia de las profesiones que impactan la necesidad del país, empresas y en la creación de nuevos negocios.

Los datos que se estudiarán para el desarrollo del modelo no infringirán la Ley de Protección de Datos porque no comprometen la identidad e integridad de los beneficiarios susceptibles de adquirir o que adquirieron un crédito educativo del ICETEX.

### 5.8. Fase 5: Implementación de *Big Data*

En esta fase se tiene en cuenta las necesidades de usuarios, que en este caso son las áreas que generan los reportes y entregan resultados según las estrategias que utilizan para mejorar metas anuales. Por ejemplo, la Vicepresidencia de crédito asigna los recursos a usuarios con pocos recursos; el área de cartera busca formas de facilitar pagos de los usuarios, y algunas otras trabajan en pro de entender el negocio, buscar lograr sus objetivos y a donde se quiere llegar. Por lo anterior en la fase 2 se implementará en el área seleccionada.

Una vez estén definidas las fases anteriores y el equipo de trabajo se realizarán las actividades relacionadas en esta fase donde se recomienda implementar el modelo CRISP-DM, que es un modelo de aplicación de minería de datos en una de las áreas seleccionadas y aterrizando las aplicaciones de *Big Data*.

Por otro lado, se puede observar en la figura 18, implementar *Big Data* evitaría solicitudes de casi 80 reportes mensuales, que realizan las diferentes áreas organización de datos.

Figura 24. Manejo actual de los datos



Fuente: Dirección de tecnología ICETEX

Implementar *Big Data*, analítica o minería de datos, se alinearía a las necesidades de las áreas. Asimismo, vale la pena destacar la importancia de trabajar conjuntamente con las coordinaciones de sistemas de información e infraestructura involucrando los estándares de MinTIC, por lo cual se tiene previsto utilizar los servicios que tiene contratado el ICETEX de gestión y orquestación de contenedores, aplicaciones y de servicios, para la integración de los sistemas de información que proveen soporte a las necesidades de negocio.

De igual manera y para cumplir con las metas anuales de las áreas, se trae el documento CONPES 3920 de 2018 y la ley 1955 de 2019, los cuales definen la normatividad aplicable

al marco de interoperabilidad para Gobierno Digital, en este caso el CONPES 3920 de política nacional de explotación de datos y con el que la Vicepresidencia de Operaciones y Tecnología junto con la Dirección de Tecnología de ICETEX busca implementar para cumplir con estas metas.

### 5.9 Plan Financiero

A continuación, se describe la proyección financiera, teniendo en cuenta algunos supuestos económicos basados en la inversión inicial, la financiación prevista, y como se podría presupuestar en 5 años los valores de este trabajo.

Para implementar estos servicios se realiza revisión de los aspectos macroeconómicos de acuerdo con la necesidad que tiene ICETEX, de servicios que *Big Data* para la innovación tecnológica del negocio, se realizó un análisis de los aspectos macroeconómicos.

*Tabla 20. Análisis económico del sector*

ANÁLISIS ECONÓMICO	
ÍNDICES ECONÓMICOS	CIFRAS
IPC 2017	6.77%
IPC 2018	5.8%
IPC 2019	4.1%
TRM AÑO 2016	\$ 2.743,39
TRM AÑO 2017	\$ 3.050.98
TRM AÑO 2018	\$ 2.951,32
TRM IV TRIMESTRE 2019	\$ 3,277
TRM MES febrero de 2020	\$ 3,440.96

Fuente: Elaboración dirección de tecnología

Los indicadores antes mencionados, afectan las posibles contrataciones. En cuanto al IPC se observa una leve disminución de uno punto siete (1.7) unidades en cuanto al año 2018 y 2019; con respecto a la TRM al comparar el promedio del año 2018 al 2019 se observa un aumento. Para el año 2019, se observa una variación del precio, aunque va subiendo entre los primeros diez (10) días del mes de febrero 2020. De mantenerse constante en la tasa representativa, favorecen los costos de los productos y servicios que se pueden complementar con lo contratado actualmente.

Dentro del análisis económico del sector público en los últimos años ha promovido el uso de tecnologías, estándares y normas aplicables, asegurando el manejo que requiere la información, con el fin de hacer más eficiente la utilización de los recursos físicos, humanos, financieros y demás que manejan las entidades públicas y se ha guiado para la ampliación y modernización de la infraestructura tecnológica, sistemas de información y la protección de datos, para suplir las necesidades, aportando así al desarrollo estratégico, misional y de apoyo, con el fin de alcanzar los objetivos institucionales planteados desde el interior de las áreas.

Ahora bien, existen diferentes herramientas contratadas en el ICETEX que permiten realizar los diferentes análisis que solicitan las áreas. Actualmente se realizan tableros por medio de la herramienta Power BI que hace parte de la licencia de office 365; sin embargo, esta no se utiliza a un 100% de lo que podría aplicar. Por ello, se pretende realizar con *Big Data* los análisis que el equipo de datos considere pertinentes, pues son los expertos en el tema y avalados por las altas direcciones, oficinas y vicepresidencias del ICETEX.

Por otro lado se utilizan las licencias que se tienen contratadas de Microsoft Azure para dar continuidad de los servicios de analítica que tiene el ICETEX desplegados en Azure y esto conforme a las características que utilizan las entidades públicas suministradas en Colombia compra eficiente, estas herramientas mejoran la capacidad en el uso de los datos, la toma de decisiones acertadas y oportunas impulsando el negocio y la productividad de cada una de las áreas.

Por lo anterior y teniendo en cuenta los índices económicos de los años anteriores se describe en la tabla 21 supuestos proyectados con un IPC de 5% y calcular un TRM de máximo 10%.

*Tabla 21. Porcentaje recursos*

Inversión y Necesidad	Inicio	Año 1	Año 2	Año 3	Año 4	Año 5	Vida útil	% Amortiz.
	Recursos							
Equipos Informáticos (portátiles)	7						3 a 5 años	10%
Licencia Software Office 365 (50 usuarios)	50						anual	10%
Licencia Software Azure DevOps (50 usuarios)	50						anual	10%
Coordinador de Grupo de datos	1						Contrato por año	5%
Arquitecto Datos	1						Contrato por año	5%
Científico de Datos	1						Contrato por año	5%
Analista de Datos	1						Contrato por año	5%
Desarrollador	1						Contrato por año	5%
Ingeniero de calidad	1						Contrato por año	5%
Experto en herramientas de analítica	1						Contrato por año	5%

Fuente: Elaboración propia

Teniendo en cuenta lo anterior, se muestra los gastos por personal y otros gastos relacionados con la operatividad del periodo proyectado, el cual se describe en la tabla 22.

Tabla 22. Proyección a cinco años

VALORES APROXIMADOS DE LICENCIAS Y RECURSOS						
Inversión y Necesidad	Inicio	Año 1		Año 2		
	Recursos	Vr. Unitario	Vr. Total	Vr. Unitario	Vr. Total	
Equipos Informáticos (portátiles)	7	\$ 1.500.000	\$ 10.500.000	\$ -	\$ -	
Licencia Software Office 365 (50 usuarios)	50	\$ 444.000	\$ 22.200.000	\$ 488.400	\$ 24.420.000	
Licencia Software Azure DevOps (50 usuarios)	50	\$ 2.323.200	\$ 116.160.000	\$ 2.555.520	\$ 127.776.000	
Coordinador de Grupo de datos	1	\$ 7.000.000	\$ 84.000.000	\$ 7.350.000	\$ 88.200.000	
Arquitecto Datos	1	\$ 8.500.000	\$ 102.000.000	\$ 8.925.000	\$ 107.100.000	
Científico de Datos	1	\$ 8.500.000	\$ 102.000.000	\$ 8.925.000	\$ 107.100.000	
Analista de Datos	1	\$ 5.000.000	\$ 60.000.000	\$ 5.250.000	\$ 63.000.000	
Desarrollador	1	\$ 6.500.000	\$ 78.000.000	\$ 6.825.000	\$ 81.900.000	
Ingeniero de calidad	1	\$ 5.000.000	\$ 60.000.000	\$ 5.250.000	\$ 63.000.000	
Experto en herramientas de analítica	1	\$ 5.000.000	\$ 60.000.000	\$ 5.250.000	\$ 63.000.000	
Año 3		Año 4		Año 5		
Vr. Unitario	Vr. Total	Vr. Unitario	Vr. Total	Vr. Unitario	Vr. Unitario	Vr. Total
\$ -	\$ -	\$ 2.000.000	\$ 14.000.000	\$ -	\$ -	\$ -
\$ 537.240	\$ 26.862.000	\$ 590.964	\$ 29.548.200	\$ 650.060	\$ 32.503.020	\$ 32.503.020
\$ 2.811.072	\$ 140.553.600	\$ 3.092.179	\$ 154.608.960	\$ 3.401.397	\$ 170.069.856	\$ 170.069.856
\$ 7.717.500	\$ 92.610.000	\$ 8.103.375	\$ 97.240.500	\$ 8.508.544	\$ 102.102.525	\$ 102.102.525
\$ 9.371.250	\$ 112.455.000	\$ 9.839.813	\$ 118.077.750	\$ 10.331.803	\$ 123.981.638	\$ 123.981.638
\$ 9.371.250	\$ 112.455.000	\$ 9.839.813	\$ 118.077.750	\$ 10.331.803	\$ 123.981.638	\$ 123.981.638
\$ 5.512.500	\$ 66.150.000	\$ 5.788.125	\$ 69.457.500	\$ 6.077.531	\$ 72.930.375	\$ 72.930.375
\$ 7.166.250	\$ 85.995.000	\$ 7.524.563	\$ 90.294.750	\$ 7.900.791	\$ 94.809.488	\$ 94.809.488
\$ 5.512.500	\$ 66.150.000	\$ 5.788.125	\$ 69.457.500	\$ 6.077.531	\$ 72.930.375	\$ 72.930.375
\$ 5.512.500	\$ 66.150.000	\$ 5.788.125	\$ 69.457.500	\$ 6.077.531	\$ 72.930.375	\$ 72.930.375

\$ 53.512.062	\$ 769.380.600	\$ 58.355.081	\$ 830.220.410	\$ 59.356.992	\$ 866.239.289	\$ 866.239.289
<b>Total valor proyecto a cinco años</b>			<b>\$ 3.886.196.299</b>			

Fuente: Elaboración propia

De igual manera impulsar el negocio y la productividad mediante las plataformas de soporte a DevOps, en particular integración nativa y soportada con Jenkins y Git Hub para integración e integración continua que traerá usuarios satisfechos reduciendo los costos y riesgos mejorando la operatividad. Tener una visión completa de las entidades o áreas del negocio para así cumplir con la normativa del gobierno digital que implementa las buenas prácticas de innovación, accesibilidad y normas de seguridad de los datos (integridad y confidencialidad), para los proyectos basados en información y generar valor en el ICETEX.

Es importante resaltar que para la implementación de este proyecto se quiere aprovechar el conocimiento y experticia del grupo de ingenieros que tiene la dirección de tecnología y las herramientas licenciadas y contratadas que tiene el ICETEX actualmente, por lo cual no genera gastos en pesos adicionales y si un mantenimiento de continuar con el grupo de trabajo y continuidad a las licencias.

## Conclusiones y recomendaciones

Una vez realizada la revisión bibliográfica se permitió identificar las diferentes metodologías que se pueden adoptar al interior de la organización con el uso de las diferentes ramas que presenta *Big Data*, lo que permiten a entidades como el ICETEX un alineamiento estratégico y consolidar una plataforma de almacenamiento y procesamiento de datos que se pueden utilizar en beneficio comercial.

La solución tecnológica propuesta para el alineamiento estratégico de *Big Data*, esta soportada bajo un ejercicio del estudio de las metodologías, lo que permitió crear una iniciativa que brindara beneficio no solo a la entidad sino también a los estudiantes y usuarios de los productos ICETEX.

Esta investigación permitió identificar que no solo se pueden utilizar los datos que reposan en la entidad y de contrario poder realizar analítica con los datos de los mensajes que presentan las redes sociales y de igual manera poder adquirir e integrar los datos de otras entidades.

Se modeló una visión tecnológica en cada una de las fases de la metodología seleccionada y una proyección de buenas prácticas como por ejemplo reducir espacios en el almacenamiento de datos y los beneficios de tener servicios en la nube para implementar políticas de desarrollo de datos abiertos.

El plan estratégico y la integración de las tecnologías de la información de ICETEX se alinea con las buenas prácticas, con las normativas de Gobierno Digital y estándares basados la implementación de datos abiertos.

El marco de referencia de MinTIC para las entidades del Estado y la Guía Técnica de Información, gobierno del dato, pueden ser implementados para el cumplimiento de las normas establecidas por el Gobierno, los cuales permiten complementar conceptos y contenidos que existen en los modelos de arquitectura de buen uso de datos.

Con los resultados de los instrumentos y con la información de las bases de datos del ICETEX, se diagnosticó la situación actual de la entidad y la urgencia por implementar o alinear la estrategia para dar cumplimiento a las metas de contribuir a la educación en los lugares más apartados de Colombia.

Se recomienda implementar metodología de analítica CRISP-DM en los procesos de algunas de las áreas que requieren analizar una cantidad específica de datos en un tema determinado como podrían ser la Vicepresidencia de Crédito y Cobranza, la Oficina Comercial y Mercadeo, la Oficina de Comunicaciones, para que sus divulgaciones sean más personalizadas para los usuarios de cada producto que genera inconformidad. Dado que esta metodología presenta una guía, estructurada en fases que permiten revisar parcial o totalmente cada paso y entender el mejor camino para la toma de una decisión.

## Lista de referencias

- Batra, Ron. (2015, enero). Trends in Data Management: Unlock the True Value Proposition of *Big Data*. *Profit*. <https://blogs.oracle.com/profit/unlock-the-true-value-proposition-of-big-data>
- Brownlow, J.; Zaki, M.; Neely, A. & Urmetzer, F. (2015). Data and Analytics- Data Driven Business Models: A Blueprint for innovation. *Cambridge Service Alliance*, <https://cambridgeservicealliance.eng.cam.ac.uk/resources/Downloads/Monthly%20Papers/2015MarchPaperTheDDBMInnovationBlueprint.pdf>
- Brinkhues, R. & Freitas, J. (2015). Information Management Capability as Competitive Imperfection in the Strategic Factor Market of Big Data. *Information Management Capability as Competitive Imperfection*, Puerto Rico, 1-7. [https://www.researchgate.net/publication/279288974\\_Information\\_Management\\_Capability\\_as\\_Competitive\\_Imprecision\\_in\\_the\\_Strategic\\_Factor\\_Market\\_of\\_Big\\_Data](https://www.researchgate.net/publication/279288974_Information_Management_Capability_as_Competitive_Imprecision_in_the_Strategic_Factor_Market_of_Big_Data)
- Chandarana, P. and Vijayalakshmi, M. (2014). *Big Data Analytics Frameworks*. *Proceedings of the International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA)*, Mumbai, 4-5 April 2014, 430-434.
- CEPAL. (2017). *La nueva revolución digital. De la internet del consumo a la internet de la producción*. Naciones Unidas. [https://repositorio.cepal.org/bitstream/handle/11362/38604/4/S1600780\\_es.pdf](https://repositorio.cepal.org/bitstream/handle/11362/38604/4/S1600780_es.pdf)
- Congreso de Colombia (2005, 30 de diciembre). Ley 1002 de 2005 (diciembre 30) por la cual se transforma el Instituto Colombiano de Crédito Educativo y Estudios Técnicos en el Exterior, Mariano Ospina Pérez, ICETEX, en una entidad financiera de naturaleza especial y se dictan otras disposiciones. <https://portal.ICETEX.gov.co/Portal/docs/default-source/documentos-el-ICETEX/estructura-jur%C3%ADdica/leyes/ley1002de2005.pdf?sfvrsn=2>
- Congreso de Colombia (2002, 19 de julio). Ley 749 de 2002 (diciembre 30) por la cual se organiza el servicio público de la educación superior en las modalidades de formación técnica profesional y tecnológica, y se dictan otras disposiciones. <https://www.ulibertadores.edu.co/images/documentos-institucionales/normatividad-externa/normatividad-educacion-superior.pdf>
- CONPES 3920. (2018, 17 de abril ) (s.f.). Política nacional de explotación de datos (Big Data). <https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3920.pdf>

Cortés, R. (2016, 26 de abril). Decideo con los grandes nombres del *Big Data* en España. Decideo. [https://www.decideo.com/Decideo-con-los-grandes-nombres-del-Big-Data-en-Espana\\_a1349.html](https://www.decideo.com/Decideo-con-los-grandes-nombres-del-Big-Data-en-Espana_a1349.html)

Departamento Administrativo Nacional de Estadística [DANE]. (2018). *Saber pra decidir 2018 Boletín Nacional*. Departamento Administrativo Nacional de Estadística [DANE], <https://www.dane.gov.co/index.php/estadisticas-por-tema/educacion/boletin-sinidel>

Decideo. (2020). *Daiichi-Sankyo integra Qlik Sense para aumentar la capacidad analítica en ventas y marketing*. [https://www.decideo.com/Daiichi-Sankyo-integra-Qlik-Sense-para-aumentar-la-capacidad-analitica-en-ventas-y-marketing\\_a1577.html](https://www.decideo.com/Daiichi-Sankyo-integra-Qlik-Sense-para-aumentar-la-capacidad-analitica-en-ventas-y-marketing_a1577.html)

Deloitte. (s.f.). *Los datos son el nuevo oro. El future de los proveedores de servicios inmobiliarios..* <https://www2.deloitte.com/ec/es/pages/financial-services/articles/los-datos-son-el-nuevo-oro.html>

García, M. & Llopis, R. (2016). La encuesta. En: García, M.; Alvira, F.; Alonso, L. & Escobar, M. (coord.), *El análisis de la realidad social. Métodos y técnicas de investigación* (331-362). Alianza.

Gallardo, J. (2009). *Metodología para la definición de requisitos de proyectos de Data Mining (ER-DM)*. [Tesis doctoral]. Universidad Politécnica de Madrid, España. [http://oa.upm.es/1946/1/JOSE\\_ALBERTO\\_GALLARDO\\_ARANCIBIA.pdf](http://oa.upm.es/1946/1/JOSE_ALBERTO_GALLARDO_ARANCIBIA.pdf)

Hartmann, P.; Zaki, M., Feldmann, N. & Neely, A. (2014). Big Data for Big Business? A Taxonomy of Data-driven Business Models used by Start-up Firms. *Cambridge Service Alliance*. [https://cambridgeservicealliance.eng.cam.ac.uk/resources/Downloads/Monthly%20Papers/2014\\_March\\_DataDrivenBusinessModels.pdf](https://cambridgeservicealliance.eng.cam.ac.uk/resources/Downloads/Monthly%20Papers/2014_March_DataDrivenBusinessModels.pdf)

Heredia, R. (2017). *La Revolución Digital y el Futuro de los Servicios Financieros*. Digital Bank.

Hernández, M. (2017, 8 de octubre). 4 ejemplos de proyectos de *Big Data* que reposan sobre MongoDB. *Decideo*. [https://www.decideo.com/4-ejemplos-de-proyectos-Big-Data-que-reposan-sobre-MongoDB\\_a1859.html](https://www.decideo.com/4-ejemplos-de-proyectos-Big-Data-que-reposan-sobre-MongoDB_a1859.html)

Hernández, R.; Fernández. C. & Baptista, P. (2010). *Metodología de la investigación* (Quinta edición).

McGrawHill. [https://www.esup.edu.pe/descargas/dep\\_investigacion/Metodologia%20de%20la%20investigaci%C3%B3n%205ta%20Edici%C3%B3n.pdf](https://www.esup.edu.pe/descargas/dep_investigacion/Metodologia%20de%20la%20investigaci%C3%B3n%205ta%20Edici%C3%B3n.pdf)

ICETEX (2017). *Informe de gestión del ICETEX 2016*. Recuperado el 19 de junio de 2017 de: <https://www.ICETEX.gov.co/dnnpro5/Portals/0/Documentos/La%20Institucion/Informe-gestion-ICETEX-2016.pdf>.

ICETEX (2016). *Plan estratégico 2016 - 2020*. Recuperado el 19 de junio de 2017. [https://portal.icetex.gov.co/Portal/docs/default-source/documentos-el-ICETEX/plan-estrategico/plan\\_estrategico\\_2016-2020.pdf](https://portal.icetex.gov.co/Portal/docs/default-source/documentos-el-ICETEX/plan-estrategico/plan_estrategico_2016-2020.pdf)

ICETEX. (s.f.). ¿Quiénes somos? <https://portal.ICETEX.gov.co/Portal/Home/el-ICETEX/quienes-somos/introducci%c3%b3n>

ICETEX. (s.f.b.). Organigrama. <https://portal.icetex.gov.co/Portal/Home/el-ICETEX/estructura-organizacional/organigrama>

Katrina Sin, Loganathan Muthu (2015, julio) (s.f.). Aplicación de grandes datos en la educación datos minería y aprendizaje análisis. Volumen 5(4). <http://ictactjournals.in/ArticleDetails.aspx?id=1893>

Lakoju, M. (2017). A Strategic Approach of Value Identification for a Big Data Project [tesis doctoral]. Brunel University London, Inglaterra. <https://bura.brunel.ac.uk/bitstream/2438/15837/1/FulltextThesis.pdf%20>

Laney, D. (2001, 6 de febrero). Gestión de datos 3D: control de volumen de datos, velocidad y variedad. Gestión de datos 3D: control de volumen de datos, velocidad y variedad. Stamford: META Group Inc.

Ministerio de Tecnologías de Información y Comunicación [MINTIC ]. (2014). *Guía Técnica de Información – Gobierno del dato* [MINTIC]. [https://www.MINTIC.gov.co/arquitecturati/630/articles-9258\\_recurso\\_pdf.pdf](https://www.MINTIC.gov.co/arquitecturati/630/articles-9258_recurso_pdf.pdf)

Nadali, A.; Naghizadeh, E. & Eslami, H. (2011). *Evaluating the success Level of Data Mining Projects Based on CRISP-DM Methodology by a Fuzzy Expert System* [ponencia]. 3rd International Conference on Electronics Computer Technology, Kanyakumari, India. [https://www.researchgate.net/publication/261109414\\_Evaluating\\_the\\_success\\_level\\_of\\_data\\_mining\\_projects\\_based\\_on\\_CRISP-DM\\_methodology\\_by\\_a\\_Fuzzy\\_expert\\_system](https://www.researchgate.net/publication/261109414_Evaluating_the_success_level_of_data_mining_projects_based_on_CRISP-DM_methodology_by_a_Fuzzy_expert_system)

Oracle. (s.f.). ¿Qué es *Big Data*? Oracle. <https://www.oracle.com/co/big-data/what-is-big-data.html>

Oracle. (2016). *An Enterprise Architect's Guide to Big Data*. Oracle. <https://www.oracle.com/technetwork/topics/entarch/articles/oea-big-data-guide-1522052.pdf>

Peña, A. (2006). *Inteligencia de Negocios: Una Propuesta para su Desarrollo en las organizaciones.* Instituto Politécnico Nacional.  
<https://es.calameo.com/read/0009834562d4384832b9e>

Peralta, Manuel. Sistema de Información. 2000. Recuperado de:  
[http://www.sinnexus.com/business\\_intelligence](http://www.sinnexus.com/business_intelligence)

Portafolio. (2017). *Así les va a las empresas en Colombia que usan tecnologías de análisis de datos.* <https://www.portafolio.co/negocios/empresas/empresas-en-colombia-que-usan-tecnologia-de-analisis-de-datos-512836>

Ragin, C. (2007). *La construcción de la investigación social. Introducción a los métodos y su diversidad.* Siglo del hombre. [http://metodos-avanzados sociales.uba.ar/wp-content/uploads/sites/216/2014/04/Investigacion\\_ragin.pdf](http://metodos-avanzados sociales.uba.ar/wp-content/uploads/sites/216/2014/04/Investigacion_ragin.pdf)

Rometty, G. (2018, 21 de enero), We need a new era of data responsibility. *World Economic Forum.* <https://www.weforum.org/agenda/2018/01/new-era-data-responsibility/>

## **Anexo 1. Preguntas – encuesta**

### **Segmento 1 – Visión estratégica, contexto y alcance del negocio**

1. ¿Cree usted que el uso de los datos que reposan en el ICETEX puede ser utilizado en el plan estratégico?
  - a. Sí
  - b. No
  
2. ¿Las IES o demás constituyentes que tienen convenio con el ICETEX podrían aportar información que contribuya a la innovación o cambio de estrategia institucional?
  - a. Sí
  - b. No
  
3. ¿Considera usted que la información de las BD del ICETEX es consistente en todos los aplicativos?
  - a. Sí
  - b. No

### **Segmento 2 – Hacia dónde dirigir la implementación del buen uso de los datos**

4. ¿Cree usted que los datos del ICETEX se pueden procesar para proponer una innovación tecnológica?
  - a. Sí
  - b. No
  
5. ¿Usted cree que la información de las bases de datos de los funcionarios del ICETEX (profesión, preparación o experiencia) contribuyen a las labores que ejercen en sus puestos de trabajo?
  - a. Sí
  - b. No
  
6. ¿Cree usted que se podría reubicar las áreas del ICETEX según las profesiones y experiencia laboral del universo de los funcionarios?

- a. Sí
- b. No

### **Segmento 3 – Estado actual de los datos y posibles usos**

- 7. ¿Usted cree que la información de las bases de datos del ICETEX podría contribuir en crear patrones sobre las posibles profesiones de los usuarios y como aportar a la empleabilidad el país?
  - a. Sí
  - b. No
- 8. ¿Usted como ingeniero podría evaluar el estado actual de las BD del ICETEX?
  - a. Sí
  - b. No
- 9. ¿Si usted tuviera una iniciativa para utilizar las BD del ICETEX en una posible modernización, cree usted que podría acceder a esa información?
  - a. Sí
  - b. No

### **Segmento 4 – Aplicación y visión futura del procesamiento de datos**

- 10. ¿Cree usted que las diferentes entidades del estado y el ICETEX podrían integrar sus bases de datos y contribuir a la modernización del país?
  - a. Sí
  - b. No
- 11. De acuerdo con los procedimientos que existen en el ICETEX, ¿usted cree que son un impedimento para el trabajo diario?
  - a. Sí
  - b. No
- 12. ¿Con el personal activo, incluido contratistas, podría usted formar un grupo que gestione y organice la aplicación de las bases de datos del ICETEX?
  - a. Sí

- b. No

**Segmento 5- Como monetizar las BD y que barreras para lograr alinearlas al plan estratégico del ICETEX**

13. ¿Conoce alguna herramienta que pueda analizar los datos históricos del ICETEX para describir o predecir una posible estrategia institucional?

- a. Sí
- b. No

Si su respuesta fue positiva podría mencionar la herramienta

---

14. ¿Cree usted que las BD del ICETEX son útiles o podrían ser utilizadas en otras entidades?

- a. Sí
- b. No

¿Podría ampliar su respuesta o nombrar qué entidad?:

---

15. Dado que el ICETEX es una entidad de nivel Nacional, ¿cree usted todos los sectores están involucrados y ayudan al cumplimiento de los objetivos estratégicos?

- a. Sí
- b. No

## Anexo 2. Cronograma

CRONOGRAMA			
FASE 1: VISIÓN ESTRATÉGICA:	DURACIÓN	INICIO	FIN
<b>Análisis y necesidades (Visión y estrategias de la entidad)</b>	5 días	1/09/2020	7/09/2020
<b>Revisión del equipo de trabajo actual</b>	10 días	8/09/2020	21/09/2020
Revisar los perfiles del personal activo de la entidad		15/09/2020	28/09/2020
Proponer participación al personal que cumpla con el perfil	15 días	29/09/2020	20/10/2020
<b>Conformación del equipo de trabajo</b>		21/10/2020	11/11/2020
<b>Elaboración de la estrategia</b>	10 días	12/11/2020	26/11/2020
• Objetivos		12/11/2020	26/11/2020
• Herramientas para proponer		12/11/2020	26/11/2020
• Presentación		12/11/2020	26/11/2020
• Sensibilización institucional		12/11/2020	26/11/2020
<b>Priorizar o seleccionar áreas de trabajos</b>		12/11/2020	26/11/2020
<b>FASE 2: DISEÑO DE HOJA DE RUTA PARA BIG DATA</b>			
Identificación de herramientas	5 días	27/11/2020	3/12/2020
Licencias de herramientas activas	5 días	4/12/2020	11/12/2020
Inventario del estado actual	5 días	14/12/2020	18/12/2020
Desarrollo y actualización de guías y procesos	8 días	21/12/2020	31/12/2020
• Gobierno de Datos (estándares, Políticas, roles y responsabilidades)			
• Storage (Qué, cómo y cuánto se almacena)			
• Security (Privacidad, confidencialidad y acceso apropiado)			
• Quality (Definición, control y mejora en la calidad de datos)			
• Interoperability (Intercambio de información)			
Apalancamiento de bases de información a partir del uso de herramientas Office 365 e históricos.			
<b>FASE 3: GENERACIÓN DE OBJETIVOS ESTRATÉGICOS DE BIG DATA</b>			
Revisión de los objetivos estratégicos actuales	5 días	4/01/2021	8/01/2021
Seleccionar y definir los objetivos	5 días	12/01/2021	18/01/2021
Presentar propuesta de objetivos alineados a la estrategia del ICETEX	3 días	19/01/2021	21/01/2021
Proponer a las áreas objetivos alineados a las tecnologías de información	6 días	22/01/2021	29/01/2021
<b>FASE 4: IDENTIFICACIÓN DE FUENTES DE DATOS</b>			
Comprobación de los seis pasos del ciclo de vida del dato y las dimensiones de la calidad de datos, según la guía MINTIC	24 días	1/02/2021	4/03/2021
• Verificar que se haya actuado acorde al plan, identificando el uso de los datos en los directorios.	3 días	1/02/2021	3/02/2021
• Identificar metadatos, datos maestros ubicación en los sistemas de información	3 días	4/02/2021	8/02/2021
• Reducción de niveles de inconsistencias y duplicidad en los datos.	3 días	9/02/2021	11/02/2021
• Revisión de que la información está disponible y sin datos ausentes	3 días	12/02/2021	16/02/2021
• Exactitud en la información, datos correctos sin errores	3 días	17/02/2021	19/02/2021
• Conformidad de los valores de los datos con formatos uniformes en las fuentes	3 días	22/02/2021	24/02/2021
• Consistencia para proporcionar información sobre el mismo objeto de datos	3 días	25/02/2021	1/03/2021
• Integridad datos completos para una buena conectividad e integración	3 días	2/03/2021	4/03/2021
<b>FASE 5: IMPLEMENTACIÓN DE BIG DATA</b>			
Necesidades de las áreas según la identificación en la fase 2.	20 días	5/03/2021	6/04/2021
Uso y sensibilización de herramientas según identificación de fase 2	10 días	7/04/2021	20/04/2021
Divulgación de beneficios del uso e implementación de Big Data	5 días	21/04/2021	27/04/2021
Mantenimiento y seguimiento	N/A	2/04/2021	N/A