



**Sistema Analítico de Información de los Laboratorios del ICA Donde se Realizan los  
Análisis de las Muestras Relacionadas con Brucelosis Bovina**

Fabián Camilo Calderón Castro

Freddy Fabián Ramírez Medina

Especialización en Big Data, Corporación Universitaria Minuto de Dios

NRC 11-938: Proyecto de grado

Ing. Efrén Eduardo Rojas Burgos

Junio 14, 2022

## Contenido

Resumen.....	7
Planteamiento del Problema .....	9
Formulación del Problema .....	11
Objetivos de la Investigación.....	11
Objetivo General .....	11
Objetivos específicos.....	11
Justificación .....	11
Justificación teórica.....	11
Justificación metodológica.....	13
Justificación Práctica.....	14
Marco de Referencia.....	14
Marco Teórico.....	15
Marco Conceptual .....	20
Big Data.....	20
Microsoft Azure .....	20
Power BI.....	21
Data Warehouse .....	21
Laboratorio nacional de diagnóstico veterinario – LNDV .....	22
La Subgerencia de Protección Animal .....	23
Enfermedades Animales.....	23
Brucelosis bovina .....	24
Análítica de datos .....	24

## SISTEMA ANALÍTICO DE INFORMACIÓN DE LOS LABORATORIOS DEL ICA

Marco Espacial.....	25
Marco Temporal.....	25
Marco Legal .....	25
Enfoque.....	27
Alcance .....	27
Diseño metodológico .....	28
Población y muestra .....	28
Recolección de información.....	28
Instrumentos .....	28
Procedimiento.....	29
Métodos para el análisis de datos .....	29
Consideraciones éticas.....	30
Desarrollo de acciones definidas en la estrategia de analítica de datos .....	32
Modelamiento proceso de analítica de datos.....	32
Entendimiento del negocio.....	33
Identificación de los diferentes indicadores de desempeño de los laboratorios.....	33
Identificación de fuentes de información .....	33
Recolección de datos.....	34
Preparación y gestión de los datos .....	35
Análisis de los datos recolectados, minería de datos .....	36
Validación y evaluación.....	36
Actividades para evaluar los resultados: .....	36
Conclusiones .....	37

## SISTEMA ANALÍTICO DE INFORMACIÓN DE LOS LABORATORIOS DEL ICA

Acciones de la dimensión tecnológica.....	37
Tablero de mando.....	37
Conclusiones.....	54
Recomendaciones y trabajos futuros .....	55
Referencias.....	56

**Lista de figuras**

Figura 1. Diagrama de flujo del proceso de análisis de datos.....	32
Figura 2. Diagrama de proceso ETL.....	34
Figura 3. Pantalla tablero de mando en herramienta Power BI .....	38
Figura 4. Filtros.....	38
Figura 5. Muestras .....	39
Figura 6. Porcentaje de muestras .....	39
Figura 7. Motivo de la solicitud.....	40
Figura 8. Tipo de solicitud.....	40
Figura 9. Pantalla indicadores de muestras por regiones en herramienta Power BI.....	41
Figura 10. Cantidad de muestras por departamento.....	42
Figura 11. Cantidad de muestras por departamento.....	42
Figura 12. Muestras con resultados positivos por municipio .....	43
Figura 13. Cuadro de información resultados positivos .....	44
Figura 14. Pantalla desempeño de laboratorios por periodos de tiempo en herramienta Power BI. .....	45
Figura 15. Desempeño de laboratorios por trimestre.....	45
Figura 16. Porcentaje del desempeño de los laboratorios por trimestre .....	46
Figura 17. Descripción de muestras realizadas de los laboratorios por mes .....	46
Figura 18. Indicador de muestras por sexo y edad del animal en herramienta Power BI.....	47
Figura 19. Descripción de resultados por edad y meses .....	47
Figura 20. Total, de cantidad de muestras y resultados positivos en los de los animales de acuerdo a su sexo .....	48

Figura 21. Total de cantidad de muestras y resultados positivos en los de los animales de acuerdo a su edad.....	49
Figura 22. Indicador de muestras por año y meses para la gestión de organizacional en herramienta Power BI. ....	49
Figura 23. Muestras procesadas por año.....	50
Figura 24. Muestras procesadas por año.....	50
Figura 25. Muestras procesadas por departamento.....	51
Figura 26. Indicador de origen de solicitud de los resultados positivos por año y meses en herramienta Power BI.....	51
Figura 27. Resultados positivos por año y mes.....	52
Figura 28. Resultados positivos por mes .....	52
Figura 29. Motivo de las muestras positivas por mes.....	53
Figura 30. Indicador de gestión de tiempos de los laboratorios y analistas en la herramienta Power BI. ....	53

## Resumen

Big data hoy en día es una preferencia a nivel global, ha evolucionado en un conjunto de ciencias, técnicas y tecnologías aplicadas que coaccionan entre sí, tomando un papel bastante importante y se han convertido en un factor imprescindible para cualquier organización que quiera ser competitiva, ayudando a la toma de acciones en el momento preciso; nosotros en este trabajo planteamos un modelo de arquitectura de datos, el cual comprende este conjunto de conceptos, los cuales inician desde el proceso de obtención de datos, captura, minería, transformación, cargue, almacenamiento, un modelado con análisis y finalmente su respectiva visualización.

De acuerdo con lo anterior y el objetivo planteado por el Instituto Colombiano Agropecuario - ICA, el presente documento nos permite obtener una exploración y una perspectiva con un análisis cuantitativo de los datos generados por los diferentes laboratorios veterinarios ubicados a nivel nacional, y es allí donde podemos identificar características, fortalezas, debilidades y oportunidades, obteniendo como resultados una síntesis y conclusiones de los procesos propios de esta actividad y de la ganadería. Por otra parte, cabe resaltar que es una temática de vital importancia para la sanidad humana como es el control de la enfermedad de la Brucelosis Bovina en Colombia.

**Palabras Claves:** *Análisis cuantitativo, Indicadores de desempeño, Estadística, Conjunto de Datos, Banco de datos.*

### **Abstract**

Big data today is a global preference, it has evolved into a set of sciences, techniques and applied technologies that coalesce with each other, taking on a very important role and have become an essential factor for any organization that wants to be competitive, and so be able to take actions at the right time with efficient strategic, tactical and operational decisions; In this work we propose a data architecture model, which includes this set of concepts, which start from the process of obtaining data, capture, mining, transformation, loading, storage, modeling with analysis and finally their respective visualization.

In accordance with the above and the objective by the Colombian Agricultural Institute ICA, this document allows us to obtain an exploration and a perspective with a scientometric analysis of the data generated by the different veterinary laboratories located nationwide, and it is there where we can identify characteristics, strengths, weaknesses and opportunities, obtaining as results a synthesis and conclusions of the processes of this activity and livestock. On the other hand, it should be noted that it is a topic of vital importance for human health, such as the control of Bovine Brucellosis disease in Colombia.

**Keywords:** *Scientometric Analysis, Performance Indicators, Statistics, Data Set, Data warehouse.*

## **Sistema Analítico de Información de los Laboratorios del ICA Donde se Realizan los Análisis de las Muestras Relacionadas con la Brucelosis Bovina**

### **Planteamiento del Problema**

El Instituto Colombiano Agropecuario (ICA), cuenta con 6 subgerencias a nivel nacional, algunas de las subgerencias depende de otras para el funcionamiento normal de los procesos que se realizan, la subgerencia de protección animal tiene como objetivo principal establecer zonas libres de enfermedades que hacen parte del control oficial; para cumplir con este objetivo trabaja junto con la subgerencia de análisis y diagnóstico (SAD) en la toma de muestras y análisis a nivel nacional, dentro de la estructura de la subgerencia de análisis y diagnóstico (SAD), se encuentra el laboratorio nacional de diagnóstico veterinario (LNDV), este laboratorio ubicado en la ciudad de Bogotá es el referente veterinario para Colombia y los diferentes laboratorios seccionales que tiene la red de diagnóstico veterinario (RLDV), en su totalidad son 26 laboratorios seccionales ubicados estratégicamente en 22 departamentos del territorio Colombiano, (Instituto Colombiano agropecuario- ICA, s.f).

El LNDV y los laboratorios seccionales tienen en su portafolio de servicio el diagnóstico para las enfermedades de control oficial que afectan a las especies animales que son explotadas económicamente, esto como parte de campañas sanitarias, apoyo a los productores y la parte de vigilancia de importación de animales, garantizando la seguridad sanitaria para el tema de comercio nacional e internacional.

El área de Brucelosis bovina es una de las áreas que mayor volumen de muestras procesa en el año, para ayudar en la ejecución de los procesos y cumplir con la ISO 17025, la subgerencia de análisis y diagnóstico se encuentra implementado un sistema LIMS (Sistema de gestión de información de laboratorios) en el LNDV y los laboratorios seccionales de

diagnóstico veterinario. La información de las muestras y los resultados se encuentra almacenada en la base de datos del sistema LIMS, que es una base de datos estructurada y centralizada en las oficinas de la tecnología de información (OTI).

La información general relacionada con las muestras (Información del intermediario, información del solicitante, información de pago, origen de la solicitud, identificación del propietario, localización del predio, agentes para bacteriología, población y la información general) se encuentra almacenada en la tabla de solicitudes y que a su vez son un folder contenedor para tener las muestras agrupadas por los grupos nemotécnicos.

En ocasiones los laboratorios no tienen la capacidad (poca cantidad de analistas con laboratorio o deficiencia de insumos y reactivos) para procesar las muestras y se tienen que enviar a otro laboratorio que en el momento pueda cubrir con las necesidades de los análisis, agregando más tiempo en la oportunidad de entrega de los resultados por el tema del transporte.

A medida que los laboratorios seccionales son puestos en el ambiente productivo en el sistema LIMS, la cantidad de información que se maneja ha empezado a crecer exponencialmente, al tener un mayor volumen de datos para extraer y procesar se han empezado a generar problemas al momento de realizar las diferentes consultas, para la toma de decisiones por parte de los diferentes directores técnicos de las subgerencias involucradas, actualmente el ICA no cuenta con una herramienta que le permita tener una visualización del comportamiento de las muestras que se procesan, lo que indica que no se pueden definir estrategias a tiempo que permita anticiparse a los diferentes temas que pueden llegar afectar los tiempos de oportunidad de respuesta especialmente para las muestras que tienen resultados positivos y dar control de seguimiento a los animales infectados.

## **Formulación del Problema**

¿Cómo estructurar y diagnosticar la información de los diferentes aplicativos que maneja el Instituto Colombiano Agropecuario ICA, relacionado con los análisis de las enfermedades de prioridad nacional enfocados en el área de la Brucelosis bovina?

## **Objetivos de la Investigación**

### **Objetivo General**

Analizar los datos con que cuenta el Instituto Colombiano Agropecuario ICA, relacionados con los análisis de las enfermedades de prioridad nacional enfocados en el área de la Brucelosis Bovina.

### **Objetivos específicos**

- Diseñar un proceso ETL de la información registrada en el SISLAB
- Construir un banco de datos adaptable en el tiempo.
- Determinar un modelo de datos para la toma de decisiones.

## **Justificación**

De acuerdo con el problema planteado, nosotros como futuros especialistas en Big data, somos un factor de la ciencia de datos que nos permite adquirir la visión y la perspectiva para la consolidación de bases de datos de manera metodológica y estructurada, para poderla transformar en información que nos permita brindar un conocimiento y al mismo tiempo crear para las organizaciones ventajas competitivas para un bien común.

### **Justificación teórica**

La brucelosis bovina en Colombia es una enfermedad infectocontagiosa conocida también como aborto infeccioso. El patógeno principal para el bovino es *Brucella abortus*. En Colombia se han aislado los biovares 1, 2 y 4. Los bovinos pueden infectarse por *Brucella suis* y

*Brucella melitensis* cuando comparten el pastoreo o las instalaciones con cerdos, cabras u ovejas infectados. La infección en bovinos por especies heterólogas de *Brucella* suele ser más transitoria que por *B. abortus*, pero acarrea un grave riesgo para la salud pública, ya que las hembras pueden excretarlas por la leche y ser más patógenas para el hombre, (López, 2014).

En Colombia, las mermas económicas en bovinos no han sido calculadas en toda su extensión, sin embargo, valoraciones realizadas hacia 1993 y recientemente actualizadas, sugieren valores hasta de 40.000 millones de pesos al año (Rivera et al., 2003; ICA, 2005). Debido a las significativas pérdidas que está representa, la brucelosis Bovina es una enfermedad de notificación obligatoria en Colombia y pertenece al grupo de las enfermedades de la lista B de la Organización Mundial de Salud Animal OIE. El programa de Prevención, control y erradicación de la brucelosis bovina parte del conocimiento epidemiológico de la enfermedad, que acceda emprender acciones encaminadas a implementar programas de detección de hatos reactivos, nivel de prevalencia en hatos, focos de infección y magnitud de su diseminación. El conocimiento y comprensión del problema han sido elementos fundamentales para sensibilizar a los ganaderos en la necesidad de programas de fincas libres de brucelosis y en la implementación de programas de vacunación.

Las pruebas para diagnosticar la Brucelosis bovina en Colombia son: el cultivo bacteriológico y la tipificación, aglutinación Rosa de Bengala, FPA, Fijación del Complemento, ELISA indirecta en suero y en leche, y ELISA competitiva, las cuales se realizan en los centros de diagnóstico del ICA. Se hace la toma de muestras serológicas y de leche para certificación de fincas libres, recertificación de fincas libres, fincas en proceso de saneamiento y movilización de bovinos y otros animales susceptibles con propósito de participar en ferias comerciales y

subastas de ganado, participar en ferias de exposición, remates de ganado, movilizar intra e Inter departamentalmente y movilizar a fincas y zonas libres de brucelosis bovina, (ICA, 2020).

El Instituto Colombiano Agropecuario (ICA) a través de Resolución 00119 de enero de 2004, ordenó dos ciclos de vacunación anual obligatoria contra la brucelosis bovina en el territorio nacional, de toda hembra bovina y bufalina entre los 3 y 8 meses de edad. Las cepas vacunales actualmente aptas para ser aplicadas en el territorio nacional son la Cepa 19 y la Cepa RB51 que se utiliza en terneras entre 3 y 8 meses de edad y en revacunación de terneras a los 15 meses de edad localizadas en fincas de zonas de alta exposición de la enfermedad o que se encuentran en predios en saneamiento, (Vega, 2006, p.69).

Como uno de los dispositivos de acción para el fortalecimiento del programa y el logro del objetivo de erradicación de la enfermedad, se implementa la Certificación de Predios Libres de Brucelosis como una estrategia diseñada a nivel internacional para avanzar en el cumplimiento de los requisitos sanitarios para la aceptabilidad de animales y sus productos en los mercados mundiales. La mencionada estrategia permite además instituir una prevalencia serológica en el país a través de muestreos a las especies susceptibles (en edad reglamentaria) y reducir la incidencia de nuevas haciendas infectadas. Adicionalmente, conseguir el control y disminución de la prevalencia en los predios detectados positivos mediante actividades de saneamiento, (Vega, 2006).

### **Justificación metodológica**

Para alcanzar los objetivos planteados por el ICA, se ha realizado un proceso metodológico ordenado y sistematizado, se realizó una síntesis con relación al plan nacional de desarrollo y los objetivos de desarrollo sostenible, revisando la misionalidad del instituto respecto a estas dos variables.

Se acude al empleo de técnicas para la extracción de los datos del aplicativo que maneja el ICA, la base de datos centralizada del aplicativo SISLAB es la principal fuente de información que se maneja para extraer y cargar la información en el banco de datos.

La arquitectura de datos se va a construir a partir de la información que se recolecte durante el proceso de muestreo y procesamiento de las muestras en los laboratorios, desarrollando tableros de control y análisis de la información, que permita la toma de decisiones y revisiones estadísticas de cómo se comporta la enfermedad en el territorio colombiano.

### **Justificación Práctica**

El ICA en uno de sus objetivos tiene el de contribuir al desarrollo sostenido del sector, acuícola, pesquero y agropecuario, mediante la vigilancia, la prevención y control de los riesgos biológicos, sanitarios y químicos, con el fin de proteger la salud de los animales y de las personas, y de esta forma poder asegurar las condiciones de comercio y consumo.

De acuerdo con lo anterior, el ICA ofrece un amplio portafolio de servicios, para la detección de las principales enfermedades que afectan a las especies animales de interés productivo para Colombia.

Por lo tanto, si se obtiene un conocimiento con el debido soporte de los datos que se tienen del sistema analítico de información, vamos a contribuir exitosamente no solo con los intereses de ICA si no también con la sanidad y el bienestar humano, junto con un desarrollo óptimo comercial.

### **Marco de Referencia**

Big Data se ha transformado en una tendencia a nivel global y aunque aún no cuenta con un concepto científico consensuado, se pronostica cada día un mayor desarrollo del mercado que lo rodea y de las áreas de investigación mancomunadas. Las técnicas y tecnologías asociadas a

Big Data implican principalmente: la captura, el procesamiento, el análisis, minería de datos y su visualización.

### **Marco Teórico**

Big Data puede ser estimada como una directriz en el avance de la tecnología que ha abierto la puerta a una nueva orientación para la comprensión y la toma de decisiones, que se utiliza para describir las grandes cantidades de datos (estructurados, no estructurados y semi-estructurados) que sería demasiado extenso y caro para cargar en una base de datos relacional para su análisis. Así, el concepto de Big Data se emplea a toda la información que no puede ser procesada o analizada utilizando herramientas o procesos tradicionales.

De acuerdo a lo anterior, Big Data y los procesos de dicha tecnología emergente simbolizan un amplio espectro de aplicaciones potenciales. El mayor reto para la inversión en Big Data se produce con relación a los proyectos vinculados a la toma de decisiones sobre un volumen amplio de datos, (Puyol, 2014).

La finalidad de Big Data consiste en capturar, almacenar, buscar, compartir y agregar valor a la variedad de datos que se pueden lograr. No siendo relevante el volumen de datos o su ambiente. Lo que interesa es su valor potencial, que sólo las nuevas tecnologías especializadas como el Big Data pueden explorar. En última instancia, el objetivo de esta tecnología es aportar y descubrir un conocimiento oculto a partir de grandes volúmenes de datos, (Puyol, 2014, p.471).

Es importante comprender que los datos estructurados que provienen de fuentes de información establecidas también se encuentra los datos no estructurados: los que llegan de las cámaras de los móviles, la Web, redes sociales, vídeos, documentos (PDF, Word), archivos de texto, edificios y sensores de las ciudades, entre otros. La diversidad de su origen, además de la

rapidez con la que se desarrolla su volumen, son algunos de los elementos que habían obstaculizado su análisis hasta ahora. El nuevo software y los nuevos modelos permiten la inscripción a los estudios tanto de un tipo como de otro.

Asimismo, se hace relevante pronosticar un comportamiento de un futuro, basándose en los datos históricos que se muestran con un mismo perfil, para esto debemos tener herramientas comerciales de Data Mining que existen actualmente. Las hay orientadas al estudio de ambientes web o al análisis de documentos de clientes y otras que son de uso más general. Su elección depende de la necesidad de la compañía y de los objetivos a corto y largo plazo que se pretendan alcanzar. La decisión de una solución de Data Mining no es una labor sencilla. Es preciso examinar junto a expertos la selección de la más adecuada para el problema de los entes empresariales. Por ello, la minería de datos se presenta como una tecnología emergente, con diversas ventajas: por un lado, resulta un buen punto de encuentro entre los investigadores y las personas de negocios; por otro, ahorra grandes cantidades de dinero a una empresa y abre nuevas ocasiones de negocios. Además, no hay duda de que trabajar con esta tecnología implica cuidar un sinnúmero de detalles debido a que el producto final involucra «toma de decisiones».

Big Data se determina por tener cinco dimensiones: volumen, variedad, velocidad, valor y veracidad.

- **Volumen:** Es la necesidad de procesamiento eficaz y complejo de subconjuntos de datos de gran dimensión que contienen información de valor para una empresa mediante tecnologías de Big Data. Es decir, la cuantía de datos,
- **Variedad:** Se define como la existencia de diversos tipos y fuentes de datos. La variedad tiene que ver con gestionar la complejidad de múltiples características de datos, incluidos los datos estructurados, semiestructurados y no estructurados. Las

organizaciones necesitan integrar y analizar datos de un complejo abanico de fuentes de información tanto de dentro como fuera de la empresa,

- **Velocidad:** Los ciclos de los negocios hoy en día prosperan a pasos acelerados y de cambios de constantes, todos los datos de una empresa tienen la misma urgencia de análisis asociada, ello lleva consigo la dimensión de los datos en movimiento desde su creación, su proceso y su análisis: por lo tanto contribuir con una mayor velocidad es primordial; La velocidad afecta claramente a la latencia: el tiempo de espera entre el momento en el que se crean los datos, el momento en el que se captan y el momento en el que están accesibles. Hoy en día los datos se generan de forma continua a una velocidad a la que a los sistemas tradicionales les resulta imposible captarlos, almacenarlos y analizarlos,
- **Valor:** Hace referencia a las ventajas del uso de Big Data como también a la determinación de cada una de sus variables,
- **Veracidad:** Es el nivel de fiabilidad asociado a ciertos tipos de datos, conseguir unos datos de alta calidad es muy importante y un reto básico de Big data, pero incluso los mejores métodos de limpieza de datos no pueden descartar la imprevisibilidad inherente de algunos datos como la economía, el tiempo o las futuras decisiones de una organización, (Puyol, 2014, pp.487-488).

Podemos llegar a concluir que el Big Data es una unión de estas características que genera una oportunidad para que las organizaciones puedan lograr una ventaja altamente competitiva en el existente mercado.

De acuerdo con lo anterior, debe indicarse que adoptar la doctrina del Big Data no es una cuestión netamente tecnológica. Necesita de una visión clara de los beneficios que genera al

basar los objetivos del negocio en datos, con valor y conocimiento de tal manera que se puedan extraer integrando tanto los datos internos como externos.

El objetivo de Big Data no es centrarse en lo que sucedió en el pasado, sino en aprender de lo sucedido y anticipar posibles situaciones para el futuro, y con esto gestionar una planeación estratégica de riesgos más concreta, ya que en esta era de la digitalización, la única constante es el cambio y la mejor forma de empezar es «hacer Big Data» es ir paso a paso, y al mismo tiempo ir tomando proyectos con objetivos bien definidos. Sin dejar de la mano a la innovación, la cual va a tener soporte de acuerdo con la información lograda.

De acuerdo con los tipos de datos que se tratan un proyecto de Big Data, podemos establecer la siguiente clasificación:

**Web and Social Media:** hace referencia al contenido web e información que es derivada de las redes sociales como Twitter, Facebook, blogs, LinkedIn etc,

**Machine-to-Machine (M2M):** M2M Hace referencia a las tecnologías que permiten conectarse a otros dispositivos. M2M utiliza dispositivos como sensores o medidores que capturan algún evento en particular (temperatura, velocidad, variables meteorológicas, etc.) los cuales transmiten a través de redes alámbricas, inalámbricas o híbridas a otras aplicaciones que traducen estos eventos en información significativa.

**Big Transaction Data:** Incluye registros de facturación, en telecomunicaciones registros detallados de las llamadas (CDR), etc. Estos datos transaccionales están disponibles en formatos tanto semiestructurados como no estructurados.

**Biometrics:** Información biométrica es la que se contiene huellas digitales, reconocimiento facial, escaneo de la retina, etc. Por ejemplo, en el área de la seguridad los datos biométricos han sido información relevante para las agencias de investigación.

**Human Generated:** Las personas crean diversas cantidades de datos como la información que guarda un Call center en sus llamadas telefónicas, correos electrónicos, documentos digitales, registros de estudios médicos, etc. (Puyol, 2014, p.498)

Los grandes volúmenes de datos (Data WareHouse) no aportará un valor añadido sin su contextualización. Es el contexto que toma protagonismo en la estrategia y que generan cambios competitivos tanto del negocio como del mercado. Las organizaciones estarán mejor equipadas para tomar ventaja de los datos los cuales pueden estar conectados a los orígenes de los mismos a través de aplicaciones on-line (Cloud).

En el proceso de ejecución de Big Data se dan una serie de principios de acuerdo a los siguientes ítems:

**Procesamiento:** Uso de tecnologías de paralelismo para que el procesamiento de los datos sea más entendible y la consiguiente determinación de los resultados.

**Recolección:** Obtención de datos de diversas fuentes.

**Dirección de los datos:** Son una determinación técnica y legislativa de los datos, su transformación, así como las labores correspondientes a las funciones de seguridad y auditoría.

**Almacenamiento:** Hace referencia a las soluciones de Storage.

**Medición:** Que consiste en la integración de correlaciones entre datos, basados en métricas inspiradas en el propio negocio.

**Gobernación:** Envuelve la facultad de disposición o gobierno de datos desde la perspectiva de negocio. (Puyol, 2014, p.504-505)

La generación de datos sigue y seguirá creciendo exponencialmente y las personas que analizan las organizaciones se están dando cuenta de que no es sólo un método escalable para

hacer Data Warehousing y Business Intelligence (BI), sino que es una tendencia altamente competitiva.

## **Marco Conceptual**

### ***Big Data***

Una arquitectura de Big data debe considerar la integración de las nuevas tecnologías y herramientas de los grandes volúmenes de datos y su integración con los datos tradicionales (bases de datos estructuradas y heredadas), al igual que la integración con la infraestructura que actualmente existe en las diferentes organizaciones y empresas.

En la actualidad, las tecnologías y herramientas de Big Data se deben centrar en la integración de diferentes tipos de datos, así como la integración de los datos que se tienen almacenados en las bases de datos estructuradas como con los datos no estructurados que se tienen en las bases de datos NoSQL.

Toda esta arquitectura de Big Data requiere de plataformas que gestionen estos grandes datos para que las organizaciones y empresas puedan obtener el máximo rendimiento. Para ello, se requieren de proveedores de soluciones que hoy día son muy numerosos y que pueden ser agrupados en proveedores de código propietario o de código abierto (open source).

### ***Microsoft Azure***

Azure fue anunciado por primera vez en el año 2008 y solo fue publicado hasta el 2010 bajo el nombre de Windows Azure, posteriormente fue renombrado como Microsoft Azure en el 2014. Azure surge para como una plataforma de cloud computing diseñada para desarrollar, crear y administrar aplicaciones y servicios a través de una red global de varios centros de datos administrados únicamente por Microsoft. Estas centrales de datos se encuentran repartidas en

todo el mundo, lo que también permite impulsar la creación de redes de trabajo internacionales para las compañías que cuentan con sedes en varios países del mundo.

En sus inicios Azure abarcaba muchos menos servicios de los que actualmente ofrece y soportaba contaba con número reducido de lenguajes de programación, software y herramientas, de Microsoft como de terceros. Esto fue cambiando progresivamente, hasta que se pudo incluir integraciones de productos propios de Microsoft como: Office 365, SharePoint, etc. Al igual que software de otras compañías: Oracle, SAP, etc. Sin embargo, la cloud computing de Microsoft comenzó ofreciendo únicamente servicios de plataforma por lo que los servicios de infraestructura no están muy definidos, comparada con otros proveedores con experiencia mucho más amplia como Amazon Web Services, (TIC Portal, 2015).

### ***Power BI***

Es un conjunto de herramientas que nos permite acceder a nuestros datos de forma segura y rápida, que nos brinda grandes beneficios para nosotros y para nuestra empresa. Es un sistema inteligente, predictivo y de un gran apoyo, es capaz de traducir los datos (simples o complejos) en gráficas que para el usuario final son muy fácil de interpretar, paneles y/o informes, además de contar con la integración de Power Query: que es el motor de extracción, transformación y carga de datos (proceso conocido como ETL) que está incluido en Excel. (PowerData, 2021)

### ***Data Warehouse***

Un Data Warehouse (Almacén de datos) es un repositorio unificado para almacenar la información de los datos recolectados de las diferentes fuentes de información de una compañía. Este tipo de repositorio puede ser físico o lógico y su principal objetivo es la captura de una gran variedad de datos para poder realizar análisis y tener acceso a la información.

En la actualidad los Data Warehouse se aloja en los servidores corporativos y con el tiempo se han empezado a migrar a servidores en la nube.

Los datos almacenados en el Data Warehouse nos permite diseñar aplicaciones de procesamiento de transacciones Online (OLTP), permitiendo a los analistas contar con información de un gran volumen, variedad y con información verídica, (Rochina, 2017).

### ***Laboratorio nacional de diagnóstico veterinario – LNDV***

El Laboratorio Nacional de Diagnóstico Veterinario (LNDV) es el laboratorio referente para la parte veterinaria de Colombia y los 26 laboratorios de la red de diagnóstico veterinario del ICA. En su portafolio de servicios tiene técnicas analíticas para el proceso de las muestras que son tomadas para las realizar control en las enfermedades que afectan a las especies animales, enfocándose en las especies que tienen un interés productivo.

La dirección técnica de diagnóstico veterinario cuenta adicionalmente con una red de 26 laboratorios seccionales ubicados estratégicamente en diferentes ciudades del territorio colombiano, las pruebas de laboratorio que ofrece cada laboratorio seccional dependen considerablemente de la esencia productiva de la zona donde encuentran ubicados. De esta forma si en la región domina la producción porcina o bovina, los centros ofrecen pruebas para el diagnóstico correspondiente de cada especie.

Dentro del laboratorio Nacional de Diagnóstico Veterinario (LNDV) se encuentra el área para la recepción de muestras en donde se maneja una ventanilla única de trámites. En esta área los usuarios pueden registrar muestras para los casos relacionados con el diagnóstico de las enfermedades de control oficial que cuentan con proyectos de erradicación, así como también con las muestras de emergencia sanitaria por presentación de epidemias causadas por enfermedades endémicas, (ICA, s.f.).

### ***La Subgerencia de Protección Animal***

La Subgerencia de Protección Animal es la encargada de la formulación, preparación y desarrollo de políticas, planes, programas, proyectos, medidas y procedimientos que están enfocados en la protección de la sanidad animal, verifica también la calidad en la producción, en la comercialización, y el uso adecuado insumos animales velando por la inocuidad en la producción de los diferentes productos que tienen un origen animal.

Según las normas internacionales adoptadas por el gobierno nacional, la Subgerencia es la encargada de establecer cuáles son las enfermedades de control oficial y de obligatoria notificación y registro; además tiene que desarrollar y administrar los diferentes programas, las campañas sanitarias para el control y la erradicación de enfermedades en el territorio Colombiano, los diferentes programas que ejecutan deben permitir llegar a zonas libres y de baja prevalencia de enfermedades, estableciendo indicadores de impacto y de mejora continua del estatus. Adicionalmente se realiza acciones preventivas junto con los productores, importadores, exportadores y las demás autoridades, para ayudar a proteger la producción animal de enfermedades y/u organismos dañinos, (ICA, s.f.).

### ***Enfermedades Animales***

Con el fin de mejorar el estatus sanitario y contribuir con la seguridad alimentaria en el territorio colombiano, el ICA define y ejecuta programas oficiales para el control y erradicación de las diferentes enfermedades endémicas de prioridad nacional que afectan a las especies animales económicamente explotables.

Del mismo modo el ICA diseña estrategias para declarar zonas libres de estas enfermedades y posteriormente mantener estas zonas libres de enfermedades que se tiene como parte del control oficial, (ICA, s.f.).

### ***Brucelosis bovina***

La brucelosis bovina es una de las enfermedades que tiene el ICA como control oficial, esta enfermedad infectocontagiosa es también conocida como aborto infeccioso. Llega afectar a bovinos de todos los géneros y edades, pero permanece con mayor frecuencia en animales sexualmente adultos, se encuentra principalmente en ganaderías de cría y leche, (ICA, s.f.).

### ***Analítica de datos***

El análisis de datos es la utilización de datos almacenados de forma estructurada con el fin de extraer conocimiento de un bloque de datos, el objetivo principal del análisis de datos es la toma decisiones de acuerdo con la información en el análisis de los datos, el concepto de analítica va más allá de esto, un concepto, tiene una composición que incluye, modelos matemáticos, estadísticos, además de metodologías que permiten un análisis de información más rápido y exacto.

Por otra parte, la analítica permite la medición de indicadores definidos, se debe tener en cuenta la frecuencia de medición de estos, es fundamental para la analítica contar con fuentes de información confiables.

El manejo y análisis de una gran cantidad de datos requiere tener en cuenta 3 características importantes, estas son:

- **Velocidad:** obtenemos esta información en tiempo real y de múltiples fuentes.
- **Volumen:** para guardar esta información requerimos servidores y otros medios más sofisticados.
- **Variedad:** obtenemos esa información en varios formatos, ya no solo en la convencional hoja de cálculo de Excel, sino en videos o audios, por ejemplo16, (Cámara de Comercio de Cali, 2022)

### **Marco Espacial**

Este proyecto investigativo se llevará a cabo principalmente en la ciudad de Bogotá, donde se encuentra ubicado actualmente El ICA en la Carrera 68A N° 24B-10 Edificio Plaza Claro - Torre 3, aunque la fuente de datos son los laboratorios se encuentra ubicados en el territorio Nacional.

### **Marco Temporal**

Este proyecto investigativo tiene una duración planificada de 11 meses, siendo esta duración el tiempo de estudio de todos los créditos de la especialización en Big Data, comprenderá el periodo septiembre 2021 – Julio 2022.

### **Marco Legal**

El ICA es una entidad pública del orden nacional, con personería jurídica, autonomía administrativa y patrimonio independiente, perteneciente al Sistema Nacional de Competitividad, Ciencia, Tecnología e Innovación, SNCCTI, y adscrita al Ministerio de Agricultura y Desarrollo Rural, se creó en 1962 mediante el decreto 1562 del 15 de junio, para coordinar e intensificar las labores de investigación, enseñanza y extensión de las ciencias agropecuarias, para el mejor y más armónico desarrollo de todas las actividades del sector y especialmente para facilitar la reforma social agraria, (ICA, s.f.).

Actualmente, el ICA tiene la responsabilidad de trabajar por el control de la sanidad agropecuaria del país aplicando las medidas sanitarias y fitosanitarias, con la orientación de acciones mediante los procesos de vigilancia epidemiológica, evaluación, gestión y comunicación del riesgo en la producción primaria. Además, lidera el desarrollo de acuerdos y negociaciones internacionales en materia fitosanitaria y zoonosanitaria para la apertura de los mercados a los productos del campo colombiano.

La sede principal de la entidad está en la ciudad de Bogotá, D.C. y cuenta con 32 gerencias seccionales, una por departamento, en los cuales tiene un recurso humano altamente calificado para atender eficientemente los problemas zoo y fitosanitarios que afectan los sistemas de producción primaria en el territorio nacional.

A lo largo y ancho del país se tienen 167 oficinas locales que trabajan por la sanidad agropecuaria y la inocuidad agroalimentaria del campo colombiano, reconocido nacional e internacionalmente por sus altos estándares como autoridad sanitaria.

El ICA en cumplimiento de lo establecido en la normatividad legal vigente aplicable a la privacidad y protección de datos personales en Colombia, actúa como responsable de los datos personales que, por virtud de sus funciones y competencias legalmente establecidas, le han sido suministradas y se encuentran en sus bases de datos siendo cada de una de las dependencias del Instituto él o la Encargado (a) del tratamiento.

Por tanto, el ICA podrá dar tratamiento a los datos personales de titulares con los cuales tiene, ha tenido o espera tener algún tipo de relación, cualquiera sea su naturaleza (civil, comercial y/o laboral, etc.) y entre los cuales se incluyen, pero sin limitarse, los grupos de interés (usuarios directos, usuarios indirectos, terceros relacionados y entidades externas)

Salvo en los casos exceptuados por la Ley 1581 de 2012 Art. 10, el ICA solicitará a más tardar en la recolección de la información, autorización del titular para capturar, almacenar, procesar, usar, circular, suprimir y en todo caso tratar los datos personales que hayan sido suministrados a la entidad por cualquier medio, bien sea digital o físico, y en desarrollo de su objeto social o con ocasión de cualquier tipo de relación civil o comercial que llegue a surgir en virtud de sus actividades conexas o propias de su naturaleza; dicha autorización deberá estar contenida en un documento físico o electrónico, (ICA, s.f.).

### **Enfoque**

En el contexto actual surgen diversas modalidades de innovación con el objetivo de tomar ventajas con los procesos y accesos a los datos y así dar solución a las problemáticas de las organizaciones. Sin embargo, para lograr esto se debe previamente poder acceder de forma abierta a los datos y a su construcción (gestión) y es aquí es dónde nosotros establecimos nuestra dirección e intervenimos con el Big data que nos ofrece una oportunidad para la mejora de los servicios de la salud animal y por consiguiente la salud de la población de consumo.

Para trascender de acuerdo con lo anterior nuestro enfoque principalmente se ha identificado en la información que genera cada uno de los laboratorios del ICA, donde obtenemos las variables cuantitativas donde se describen Numero de reporte, numero de Solicitud, fechas de recepción, fechas de la toma de la muestra, fechas de análisis, fechas de resultados. Identificación de la muestra y edades de los animales, al igual encontramos otras variables categóricas como correos de envió, Nombre del Laboratorio, Nombre del responsable, descripción de la muestra, especie, predio, municipios, departamento, raza, prueba solicitada, objeto del análisis, sexo del animal, códigos internos asignado, nombre de los propietarios de los bovinos, método de la prueba y los resultados acompañados de las observaciones.

### **Alcance**

El alcance de la investigación es descriptivo ya que con la consolidación de la información generada por los diferentes laboratorios podemos observar que las grandes bases de datos históricas albergan datos de los diferentes entes que interviene en todo el proceso y esto nos permite mediante la minería de datos llegar a brindar información valiosa para el gremio; Como la identificación de las zonas (Departamentos, Municipios, etc.) afectadas por esta enfermedad, la cantidad de animales por región, La duración de los procesos de laboratorios

desde que se genera una solicitud hasta cuando es generado el reporte de resultados. Se logra validar los métodos de pruebas, Las razas, edades y genero de los animales más afectados por la brucelosis, al igual que los desempeños de los laboratorios a nivel nacional.

## **Diseño Metodológico**

### **Población y muestra**

Colombia es uno de los países que no es la excepción con la incidencia de brucelosis de los animales en Latinoamérica. A pesar de los programas ya implementados por el ICA en la población animal y de los avances alcanzados, actualmente continúa siendo un problema de salud pública. Por lo tanto, para responder el presente trabajo tiene como objetivo evaluar el total de las muestras de los rumiantes bovinos que se diagnostican en los diferentes laboratorios a nivel nacional, esto nos permite cuantificar y operar con objetivos claros de acuerdo con las variables que se nos presentan de una manera más eficiente en el momento de presentar la información.

### **Recolección de información**

Los datos se originan en los diferentes laboratorios y se centralizarían en un sistema ya establecido por el ICA donde se recolectan para realizar el proceso ETL y su posterior almacenamiento en un Data Warehouse.

### **Instrumentos**

La información se va a extraer de la base de datos del sistema de información para laboratorios SISLAB, esta información se tiene en un motor de base de datos Oracle 11g, para la recolección de las información en los servidores de Azure se instaló el drive de Oracle y un Gateway para poder enlazar la base de datos, al momento de extraer la información se va a procesar con la herramienta synapse analytics, en la cual se va a realizar El Data Cleansing de

toda la información, antes de cargarla en el Data Warehouse destinado para los proyectos de los laboratorios del ICA.

### **Procedimiento**

Los datos que se van a extraer de las diferentes fuentes de información, van a estar almacenada en los servidores que se encuentra en el centro de datos del ICA, el acceso es muy restringido y únicamente los funcionarios que trabajan en el análisis de datos tienen acceso a esta información.

Se debe diligenciar el acuerdo de confidencialidad, que es firmado por el jefe de la OTI y el funcionario que va a tener acceso a los datos, donde se describe que la información compartida pertenece al ICA, y la misma es considerada sensible y de carácter restringido en su divulgación, manejo y utilización. Dicha información es compartida en virtud del desarrollo de las actividades que se van a realizar para crear los sistemas OLAP. Los permisos son restringidos por cada usuario para el ingreso a los servidores donde se manejan los diferentes Data Warehouse que se va a trabajar, estos servidores se encuentran en Microsoft Azure, cada usuario debe tener una cuenta para poder ingresar al servidor, esta va a contar con un usuario y una contraseña que son intransferibles.

La analítica de los datos dentro de Azure se realizará con Synapse Analytics y se configuraran los tableros de control con Power BI, para posteriormente publicar los tableros en el sitio web del ICA, donde los usuarios finales van a tener acceso a la diferente información que se tienen en el Data Warehouse.

### **Métodos para el análisis de datos**

Los resultados de las muestras se normalizarán utilizando un control de información por medio de la herramienta Synapse-analytics,

Es un servicio de análisis ilimitado que reúne la integración de datos, el almacenamiento de datos empresariales y el análisis de macrodatos. Le da la libertad de consultar los datos como quiera, usando opciones sin servidor o dedicadas, a gran escala. Azure Synapse combina estos mundos y ofrece una experiencia unificada para ingerir, explorar, preparar, transformar, administrar y servir datos con el fin de satisfacer las necesidades inmediatas de inteligencia empresarial y aprendizaje automático. (European Knowledge Center for Information Technology, 2015, p.1)

Los datos procesados y almacenados en el Data Warehouse se analizarán mediante Power BI, donde se van a utilizar objetos visuales, que se van a crear con el objetivo que los usuarios finales puedan tomar decisiones basados en el análisis de la información.

### **Consideraciones éticas**

Big Data como su nombre lo describe es una gran cantidad de datos, que normalmente se incrementan con el tiempo. El Big Data es como la memoria para el cerebro, lo que hace ésta es proporcionar respuestas a preguntas que ni siquiera sabían que las debían tener en cuenta.

Sin embargo, la evolución del Big Data ha creado en algunos países el control del gobierno sobre los ciudadanos mediante elementos tecnológicos utilizados para la predicción de crímenes, reconocimiento facial, sistemas de rastreo y sistemas de puntuación de ciudadanos, etc, que a su vez a los mismos se le otorgan beneficios o descuentos de acuerdo a su conducta, generando una perspectiva de un mundo futuro en donde la libertad del ser humano quizá puede ser permeada por el avance agigantado del Big data.

En concordancia con la declaración universal de los Derechos Humanos, (ONU) la libertad es un derecho sagrado e imprescriptible que todos los seres humanos poseen, es la facultad de obrar según la voluntad individual, respetando la ley y el derecho ajeno. Si

entendemos el concepto de libertad en la voluntad individual, puede que la misma se vea expuesta por los avances en tecnología en el marco de Big Data, al realizar un análisis en la historia y evolución de estas tecnologías podemos ver un avance de estas con una gran tendencia hacia el control de toda actividad humana, es en esta apuesta donde se podría evidenciar la sumisión en su máxima expresión.

De acuerdo con lo anterior, deja un precedente que es importante abordar, desde el surgimiento e historia del Big data, es el concepto de libertad y como la humanidad debe avanzar apoyada por la tecnología en la demanda de la transformación digital que actualmente lleva el mundo sin dejar de lado los principios éticos de la humanidad.

Es importante también resaltar que el Big Data se compone de datos que se derivan de las tecnologías modernas como: los web logs, sensores, maquinaria, vehículos, búsquedas en internet, las redes sociales, teléfonos inteligentes, registros de llamadas, dispositivos con GPS. Esto ayuda a la reducción de costos, a tomar decisiones de manera inmediata y más rápidos de lo común, y genera nuevos productos y servicios. Los datos son tan diversos y diferentes dependiendo del enfoque y los beneficios que se quiere que se sacar provecho, no se puede sectorizar o clasificar ya que es una metodología que se puede aplicar a todo.

Podemos concluir que la tecnología e incluso el surgimiento del Big Data no tiene ética, pero la humanidad o los especialistas son los responsables de abanderar su buen uso y su integro proceder de la mano de la ética con un bien común para todos.

De igual forma y de la mano con el desarrollo e implementación un sistema de Big data para el ICA, no estamos alineados únicamente con los principios éticos planteados si no que paralelamente nos regimos bajos los parámetros establecidos por la misma organización que son los establecidos por el gobierno Nacional con los objetivos de Desarrollo sostenible (ODS).

también conocidos como Objetivos Mundiales, los cuales se adoptaron por todos los Estados miembros como un llamado universal para poner fin a la pobreza, proteger el planeta y garantizar que todas las personas gocen de paz y prosperidad, (Fundación ABD, s,f).

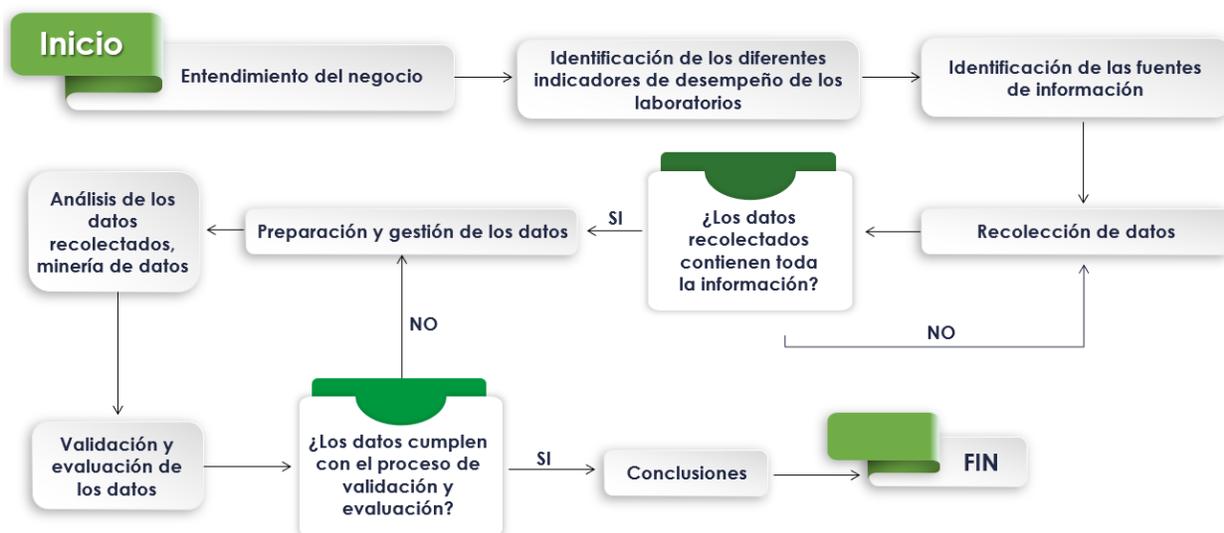
### Desarrollo de Acciones Definidas en la Estrategia de Analítica de Datos

#### Modelamiento proceso de analítica de datos

En esta parte se podrá encontrar el proceso que se trabajó para el análisis de los datos de las muestras del área de Brucelosis de los laboratorios de diagnóstico veterinario del ICA, a continuación, se puede visualizar el diagrama con el proceso con el que se trabajara en el análisis de los datos. Ver la figura 1.

#### Figura 1

*Diagrama de flujo del proceso de análisis de datos*



Nota: Elaboración propia

Para el proceso de análisis de datos, es importante tener definidas algunas etapas que nos permitan validar las conclusiones que se puedan extraer de los datos analizados; para realizar todo el proceso de análisis de los datos, es requerido cumplir con algunas etapas que contribuyan

de una forma favorable a sacar las conclusiones que se pueden extraer después de realizar el análisis de los datos.

### **Entendimiento del negocio**

Validar los objetivos del proyecto y cada uno de los requerimientos desde el núcleo del instituto, para poder convertir el conocimiento en un proceso de minería de datos, y poder tener un plan de trabajo inicial que cumpla con los objetivos. Además de poder identificar los riesgos y requerimientos generales de que se pueden tener en el instituto para el desarrollo del proyecto.

### **Identificación de los diferentes indicadores de desempeño de los laboratorios**

Identificar los principales indicadores de desempeño de los laboratorios de la red de diagnóstico veterinario, estos son útiles para la medición de los procesos y posteriormente serán la fuente de información para la toma de decisiones por parte de los directores técnicos, era muy importante identificar los tiempos de oportunidad de entrega y los tiempos analíticos de los laboratorios para el análisis y procesamiento de las muestras de brucelosis.

### **Identificación de fuentes de información**

Identificar las fuentes de información era una de las actividades fundamentales para cumplir con el objetivo del proyecto, de acuerdo a los diferentes procesos que se relacionan con la toma de muestras, el análisis y el reporte de resultados, existen actividades desde diferentes subgerencias que impactan el tiempo de respuesta oportuna para la entrega de resultados, por ejemplo el tiempo que tarda la muestra en ser transportada al laboratorio impacta fuertemente el tiempo de entrega, generando posibles contagios en otros animales si no se tiene el animal en cuarentena preventiva. En general el proceso que se analiza con mayor detenimiento es el del resultado final emitido por los analistas, las zonas de donde provienen las muestras que se encontraron positivas, la cantidad de muestras que llegan a los diferentes laboratorios,

identificando la cantidad de insumos consumidos en un intervalo de tiempo y el tiempo de entrega de los resultados, pues de allí se puede obtener los resultados de desempeño actuales de los diferentes laboratorios de la red de diagnóstico veterinario.

Se identificaron las tablas de la base de datos relacionadas con cada uno de los procesos realizados por los diferentes roles que interactúan con las muestras y módulos de SISLAB.

### Recolección de datos

Lo definido en el alcance del proyecto era consultar y extraer todos los datos almacenados en la base de datos de SISLAB, para realizar las consultas a la base de datos el personal de la OTI configuro en el entorno de trabajo Oracle SQL Developer una conexión directa a la base de datos de producción de SISLAB, en donde se creó una consulta (PL-SQL) para extraer el conjunto de datos históricos que se tenían hasta la fecha de corte de la recolección de los datos. Los registros que selecciono la consulta de la base de datos se exportaron en un documento .CSV para poder manipular la información histórica que se tiene de las muestras relacionadas con la brucelosis.

### Figura 2

*Diagrama de proceso ETL*



Nota: Elaboración propia

## **Preparación y gestión de los datos**

Los datos extraídos en el documento CSV se cargaron en el aplicativo Power BI, en donde se realizó el proceso de transformación de la data, posteriormente al tener la data limpia se volvió a cargar a un tabla creada en SISLAB que va a contener toda la información del histórico de muestras y las muestras que ingresen al sistema posteriormente a la fecha de corte de extracción de los datos, el cargue inicial de la data transformada se realizó con ayuda de Oracle SQL Developer y para la cargar de información de las muestras a futuro se creó un disparador dentro de SISLAB, este sistema maneja un código de configuración propio (LIMS Basic) pero que toma las mismas funciones de C# adicionando funciones propias de los sistemas LIMS, el disparador se ejecuta cuando los líderes de los laboratorios autorizan las muestras y generan un reporte de resultados en estado definitivo.

La sabana de datos que se configuro tiene permisos para que el usuario: analitica.bi pueda tomar la información y cargarla en los bancos de datos (Data Warehouse) configurados en los servicios en la nube de Azure, para poder realizar los respectivos análisis de datos y visualización en la herramienta Power BI, que es la herramienta que tiene la OTI para la visualización de los diferentes tableros de control.

Dentro de este proceso se identificaron los datos que son útiles para su posterior análisis, no se tuvo en cuenta toda la información extraída en el conjunto de datos para los análisis que se realizaran, pero el ICA solito cargar más datos para posteriores proyectos de Big Data, luego de definir los datos útiles el siguiente paso realizo fue la integración de los datos definidos, era muy importante que todos los datos estuvieran relacionados entre sí.

### **Análisis de los datos recolectados, minería de datos**

Después de realizar las actividades previas, fue muy importante analizar la información transformada y validar que nos indicaba el análisis de esa información, si dentro de los datos que teníamos se identifica un comportamiento que nos permitiera visualizar la situación actual del área de Brucelosis, además si se requiere extraer más datos para validar los indicadores establecidos por el laboratorio, también se debía contemplar el tema de validar si la información transformada requiere transformaciones adicionales, en esta etapa la información tenía que ser consistente.

### **Validación y evaluación**

Determinar los modelos que son más efectivos para ser considerados acorde a los criterios del núcleo del área de brucelosis, se tenía que realizar una verificación de la calidad de los diferentes comportamientos identificados, acompañado de la validación de los datos tomados y transformados, cuando la información supero esta etapa se dio visto bueno de la información que ha pasado por el proceso de ETL.

#### ***Actividades para evaluar los resultados:***

- Entender los resultados encontrados en el proceso de análisis
- Interpretar los resultados para poder ser aplicado
- Validar los resultados comparándolos con los conocimientos de los hechos que se tienen para evaluar si la información analizada es útil para la toma de decisiones por parte de los directores técnicos
- Generar recomendaciones para futuros proyectos de Big Data en el ICA
- Aprobar los modelos de análisis según los resultados encontrados en la validación

## **Conclusiones**

Al finalizar los procesos y tener los comportamientos indicados, se procede con las conclusiones que se identificaron en la realización del análisis de los datos, con el fin de mejorar el desempeño de los laboratorios del ICA que procesan las muestras de brucelosis, y la certificación y mantenimiento de las zonas libres de esta enfermedad de control oficial, aportando a la sanidad animal en el territorio colombiano.

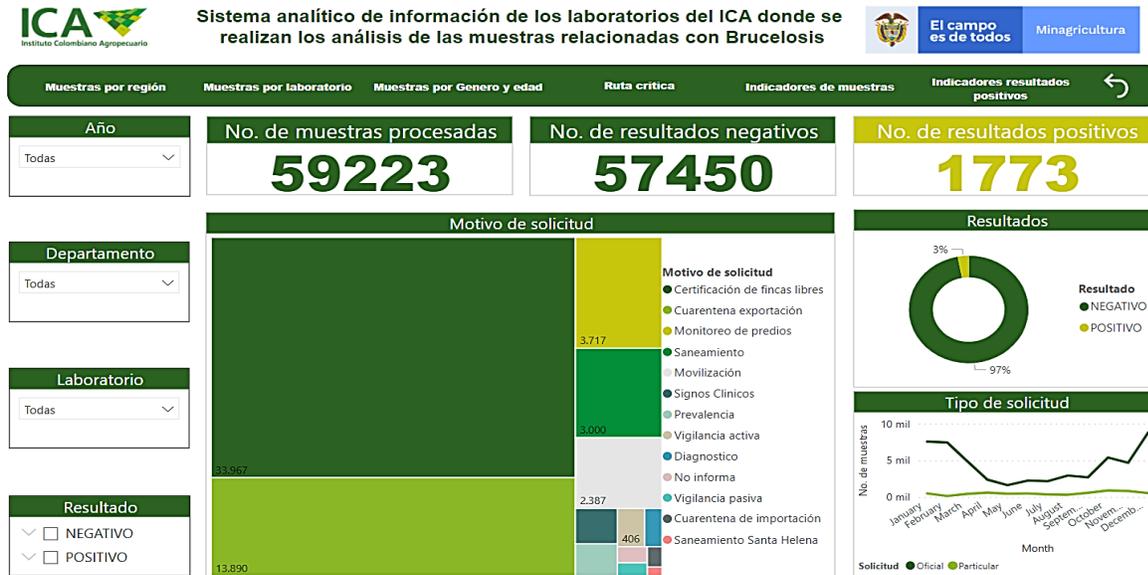
### **Acciones de la dimensión tecnológica**

#### **Tablero de mando**

Una de las labores a desarrollar, es la visualización de forma gráfica de los indicadores sugeridos en una herramienta tecnológica, el instrumento que se utilizará es Power BI, la cual admite obtener los indicadores propuestos. La herramienta permitirá la observación de los indicadores de manera general o individual, esta visualización es llamada Tablero de mando, pues concatena los diversos indicadores en una sola pantalla y permite tener una visión general. El cuadro de mando contará con filtros en donde el usuario podrá ejecutar varias selecciones como los años de consulta o el año, El departamento, los laboratorios que reportan a nivel nacional como también pueden visualizar la cantidad de muestras procesadas y de las cuales cuantas fueron negativas y positivas. En esta misma instancia podemos observar la cantidad de muestras por motivos y tipos de solicitud.

**Figura 3**

*Pantalla tablero de mando en herramienta Power BI*



Nota: Elaboración propia

A continuación, veremos los primeros indicadores y filtros de la fuente de información que fueron las bases de datos comprendidas entre el año 2017 al 15 de abril de 2022.

**Figura 4**

*Filtros*



Nota: Elaboración propia

Al ejecutar estos filtros podremos obtener los principales parámetros de indicadores que vamos a analizar teniendo como los periodos de tiempo por año, departamento, Laboratorio y los resultados de las muestras.

### Figura 5

#### Muestras



Nota: Elaboración propia

En esta tarjeta de visualización se determina el indicador de la cantidad total de muestras realizadas en los bovinos y cuantas de ellas los resultados fueron negativos y positivos.

### Figura 6

#### Porcentaje de muestras

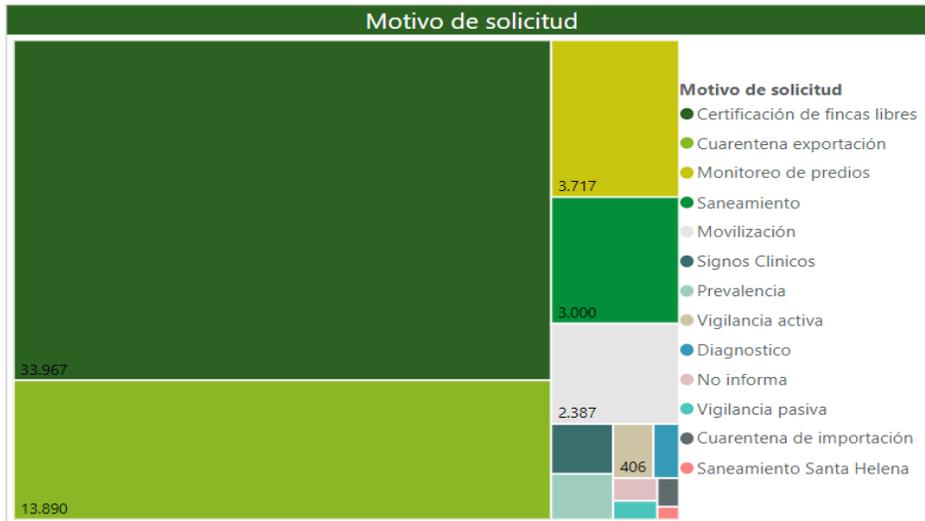


Nota: Elaboración propia

Esta ilustración brinda el porcentaje de los animales que se encontraron contagiados y los que no.

**Figura 7**

*Motivo de la solicitud*

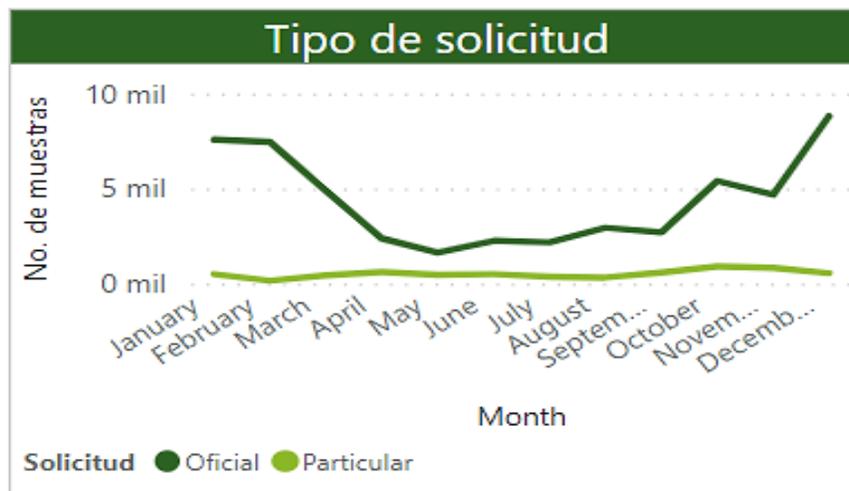


Nota: Elaboración propia

En el diagrama de motivo de la solicitud se encontrará la proporción de las cantidades de las muestras solicitadas y como lo menciona su nombre se determina el concepto de análisis o por qué se llevó a cabo.

**Figura 8**

*Tipo de solicitud*

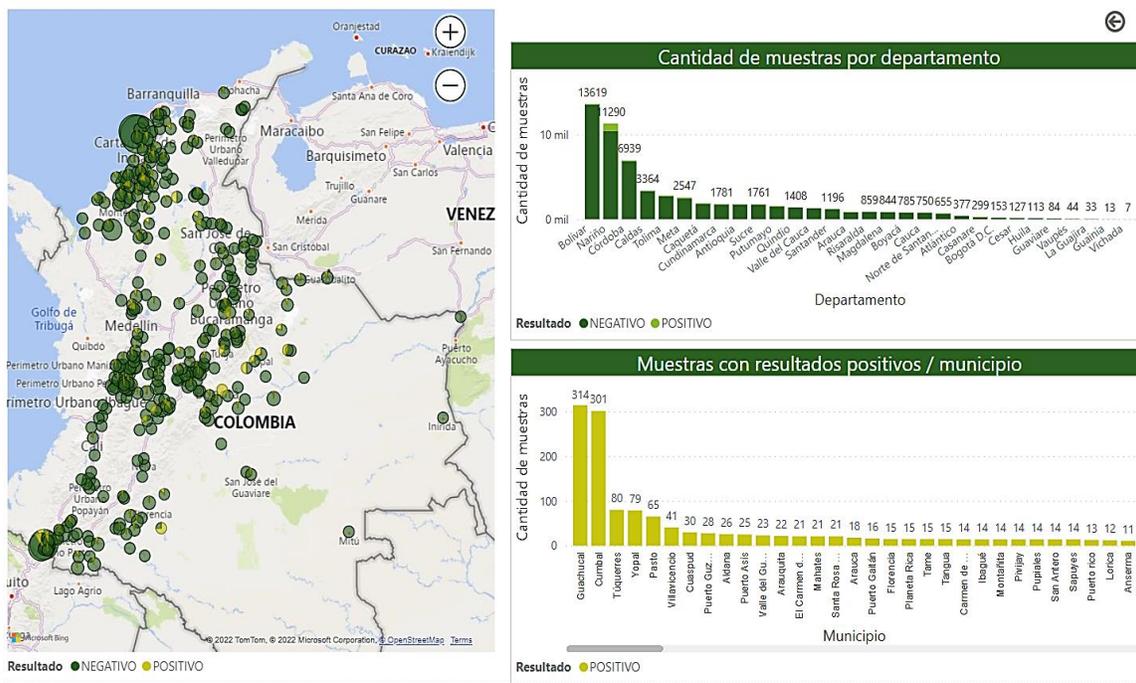


Nota: Elaboración propia

En esta figura tenemos el número de muestras por la línea de tiempo en meses donde se puede observar la tendencia de la cantidad de muestras que se realizaron oficialmente por ICA o si las solicitudes se realizaron de manera particular.

**Figura 9**

*Pantalla indicadores de muestras por regiones en herramienta Power BI*



Nota: Elaboración propia

En un segundo tablero se encuentra de manera gráfica la determinación de resultados de las muestras de los diferentes laboratorios, los cuales podemos ubicar geográficamente en el mapa de Colombia y visualizar indicadores por departamento y municipio.

**Figura 10**

*Cantidad de muestras por departamento*



Nota: Elaboración propia

Esta gráfico brinda la visualización geográfica donde se han realizado las pruebas de laboratorio dando a conocer las zonas del país libres de brucelosis.

**Figura 11**

*Cantidad de muestras por departamento*



Nota: Elaboración propia

En este ranking se ilustra los departamentos con la cantidad mayor a la cantidad menor de las muestras tanto positivas y negativas ofreciendo quien de ellos obtuvo un mayor protagonismo.

**Figura 12**

*Muestras con resultados positivos por municipio*



Nota: Elaboración propia

En esta figura observamos también un listado de mayor a menor de los casos positivos que se presentaron por municipio y al ubicar el cursor del mouse sobre cada una de las barras de la gráfica se desplegará una ventana donde se puede observar el detalle de los casos positivos como el número de la solicitud, la fecha, la cantidad de muestras y el promedio de la edad de los animales.

**Figura 13***Cuadro de información resultados positivos*

**Departamento**  
Antioquia

**Municipio**  
Abejorral

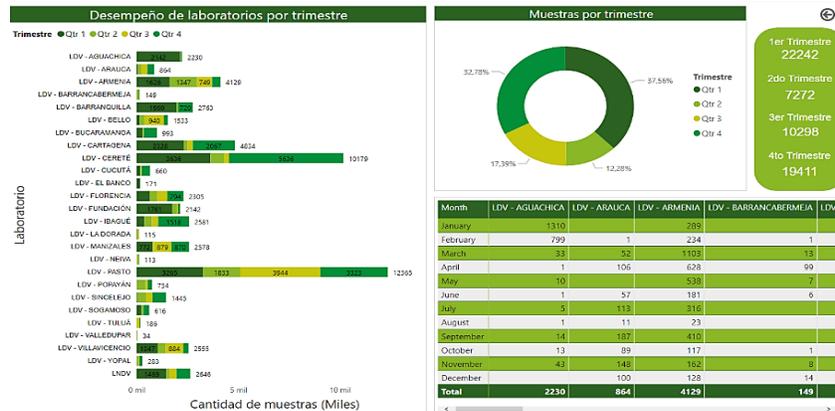
No. de solicitud	Year	Month	Day	Cantidad de muestras	Promedio de edad (Meses)
S01170000005	2017	January	18	19	57
S01170000010	2017	January	24	3	47
S01190000041	2019	January	9	3	36
S01210000003	2021	January	6	187	64
S01220000041	2022	January	11	57	83
S01220000230	2022	January	31	25	50
S03220000004	2022	January	14	800	24
S03220000005	2022	January	24	10	48
S03220000007	2022	January	25	500	24
S05190000024	2019	January	17	40	65
S05190000025	2019	January	17	66	77
S05190000029	2019	January	18	2	30
S05190000036	2019	January	21	2	60
S05200000008	2020	January	9	54	77
S05200000028	2020	January	13	70	74
S05200000089	2020	January	29	32	30
S05200000090	2020	January	29	10	33
S05200000091	2020	January	29	13	30
S08210000004	2021	January	25	13	45
S10190000007	2019	January	15	2	25
S10190000008	2019	January	17	1	40
<b>Total</b>				<b>59223</b>	<b>51</b>

Nota: Elaboración propia

Para tener un control de las solicitudes que tienen muestras con resultados positivos, se configuro la herramienta de información para que consultara las diferentes solicitudes con los datos requeridos para que los directores puedan realizar los respectivos seguimientos

**Figura 14**

*Pantalla desempeño de laboratorios por periodos de tiempo en herramienta Power BI.*

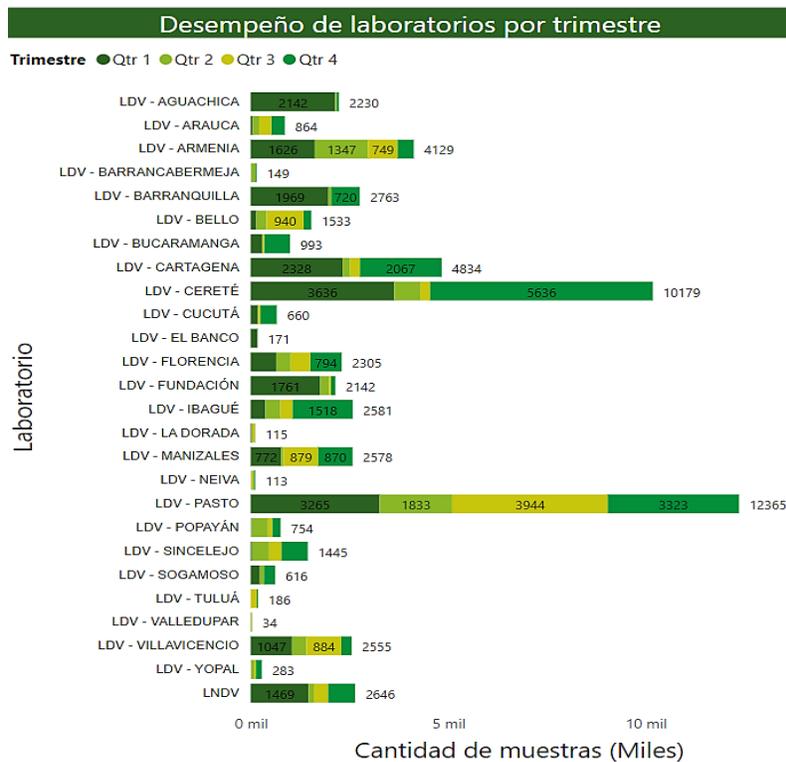


Nota: Elaboración propia

Se encontrará un indicador para los laboratorios por trimestres y meses

**Figura 15**

*Desempeño de laboratorios por trimestre*

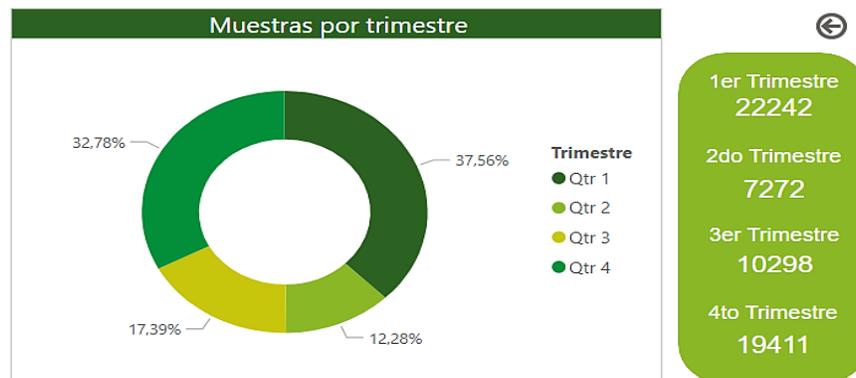


Nota: Elaboración propia

En esta figura se puede visualizar el ejercicio de cada uno de los laboratorios e indicar las cantidades que se registraron por trimestre así identificando cuales fueron los periodos con una mayor actividad.

**Figura 16**

*Porcentaje del desempeño de los laboratorios por trimestre*



Nota: Elaboración propia

Aquí podemos determinar qué cantidad y porcentaje de muestras realizaron por trimestre los diferentes laboratorios.

**Figura 17**

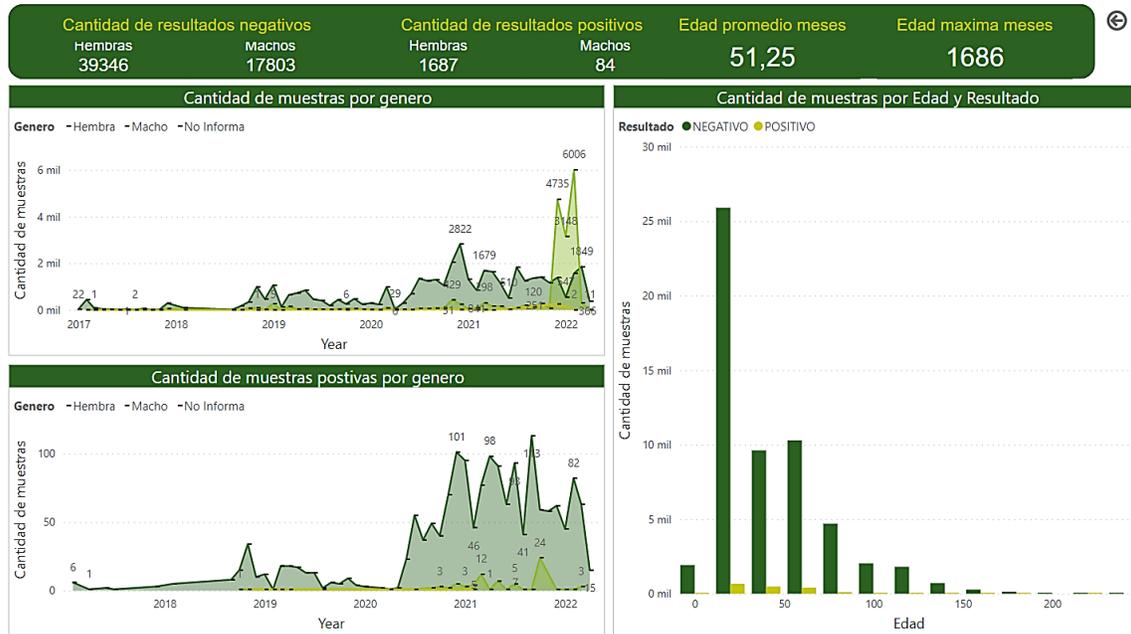
*Descripción de muestras realizadas de los laboratorios por mes*

Month	LDV - AGUACHICA	LDV - ARAUCA	LDV - ARMENIA	LDV - BARRANCABERMEJA	LDV -
January	1310		289		
February	799	1	234		1
March	33	52	1103		13
April	1	106	628		99
May	10		538		7
June	1	57	181		6
July	5	113	316		
August	1	11	23		
September	14	187	410		
October	13	89	117		1
November	43	148	162		8
December		100	128		14
<b>Total</b>	<b>2230</b>	<b>864</b>	<b>4129</b>		<b>149</b>

Nota: Elaboración propia

**Figura 18**

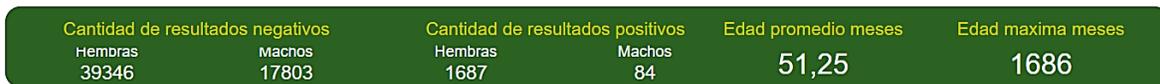
*Indicador de muestras por sexo y edad del animal en herramienta Power BI.*



Nota: Elaboración propia

**Figura 19**

*Descripción de resultados por edad y meses*

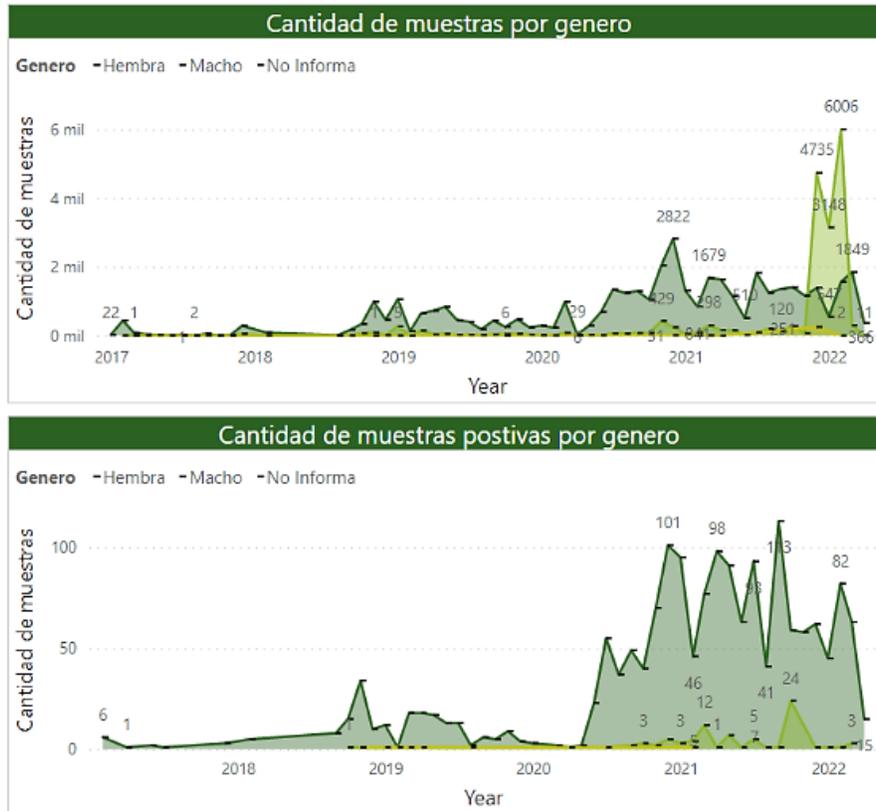


Nota: Elaboración propia

Comparativo entre los resultados positivos y negativos para establecerlos de acuerdo a su edad y sexo de los bovinos.

**Figura 20**

*Total, de cantidad de muestras y resultados positivos en los de los animales de acuerdo a su sexo*

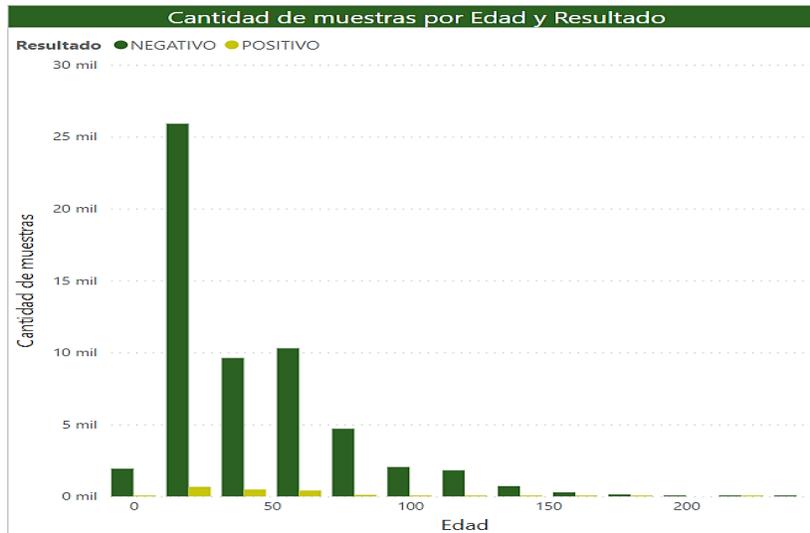


Nota: Elaboración propia

De manera gráfica se puede comparar la cantidad de muestras realizadas por año segmentándolo por sexo de los bovinos y así determinar su tendencia en el transcurrir del tiempo.

**Figura 21**

*Total de cantidad de muestras y resultados positivos en los de los animales de acuerdo a su edad*



Nota: Elaboración propia

En esta figura se establece la cantidad de casos positivos y negativos por edad del animal.

**Figura 22**

*Indicador de muestras por año y meses para la gestión de organizacional en herramienta Power BI.*

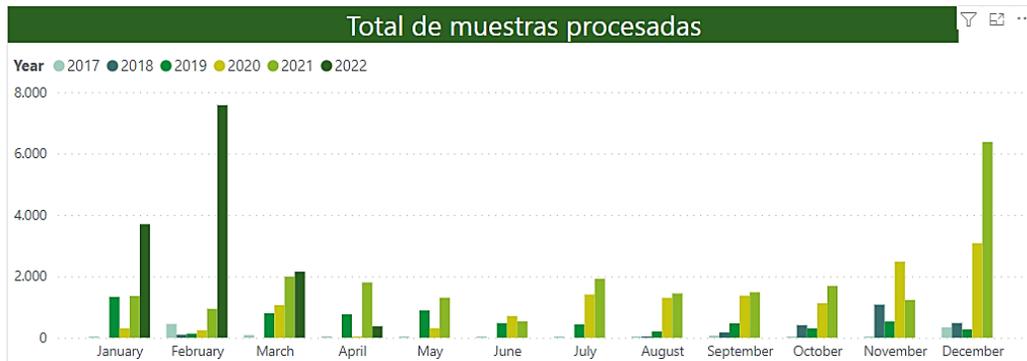
BI.



Nota: Elaboración propia

**Figura 23**

*Muestras procesadas por año*

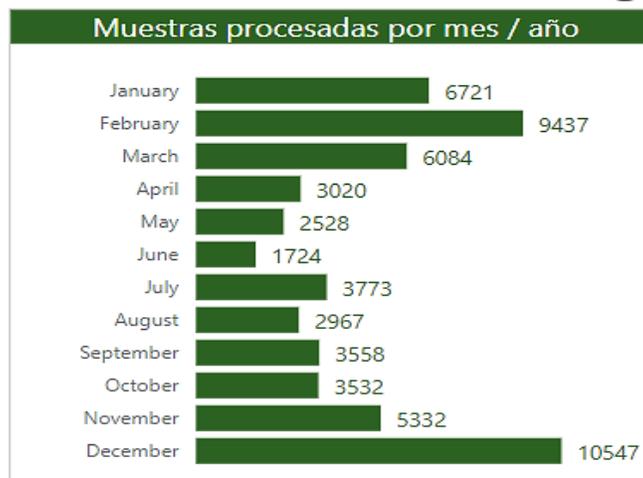


Nota: Elaboración propia

En esta gráfica el objetivo es evidenciar la cantidad de muestras de los diferentes años por meses y así indicar cuáles son sus intervalos.

**Figura 24**

*Muestras procesadas por año*

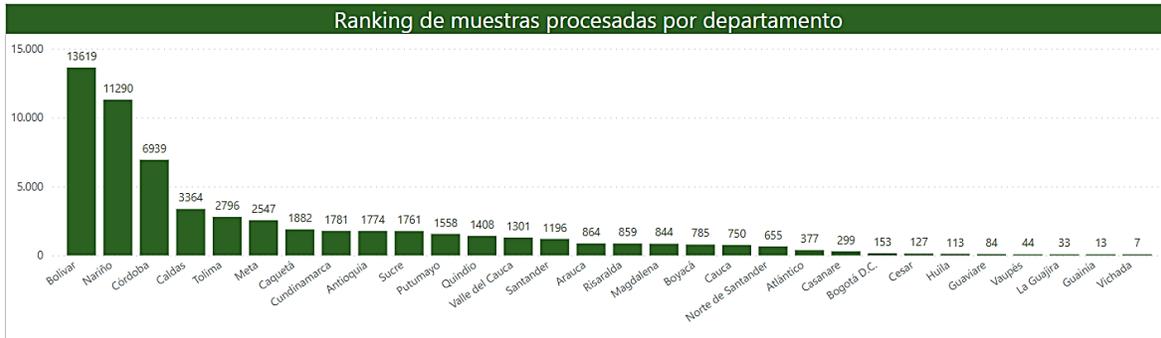


Nota: Elaboración propia

En este diagrama facilitará la visualización del mes que con mayor y menor número de muestras.

**Figura 25**

*Muestras procesadas por departamento*



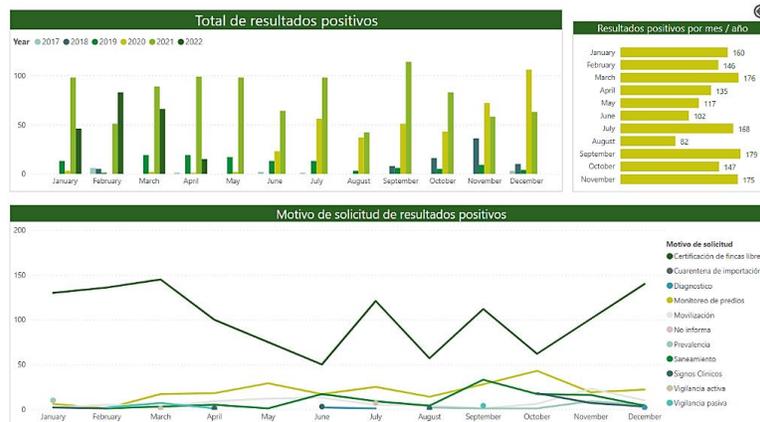
Nota: Elaboración propia

Esta ilustración nos ayuda a relacionar el departamento o provincia con la cantidad de muestras a través del tiempo.

**Figura 26**

*Indicador de origen de solicitud de los resultados positivos por año y meses en herramienta*

*Power BI*



Nota: Elaboración propia

En esta gráfica se puede observar el desempeño de las muestras positivas por año y meses con el objetivo de determinar el origen de la solicitud.

**Figura 27**

*Resultados positivos por año y mes*



Nota: Elaboración propia

En esta figura se puede determinar los meses en transcurso de los años y meses en donde se presentan el mayor y menor número de casos positivos para el debido seguimiento.

**Figura 28**

*Resultados positivos por mes*

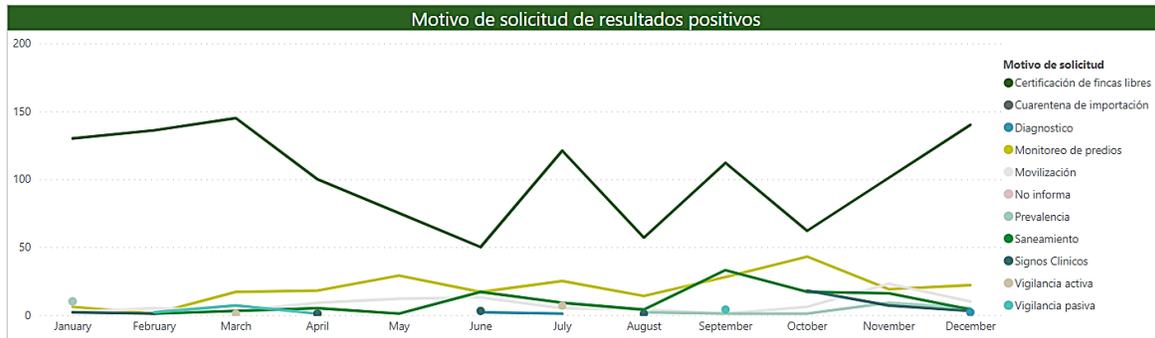


Nota: Elaboración propia

En esta figura se establece los meses donde se han presentado el mayor número de casos positivos.

**Figura 29**

*Motivo de las muestras positivas por mes*



Nota: Elaboración propia

En esta ilustración se agiliza la visualización del origen de la solicitud de los resultados positivos en el periodo de tiempo mensual.

**Figura 30**

*Indicador de gestión de tiempos de los laboratorios y analistas en la herramienta Power BI.*

Ruta crítica			
Laboratorio	No. de muestras	Ruta crítica	Cantidad de días
LDV - AGUACHICA	2230	6	3,75 ▲
LDV - ARAUCA	864	6	2,03 ▲
LDV - ARMENIA	4129	6	1,81 ▲
LDV - BARRANCABERMEJA	149	6	1,65 ▲
LDV - BARRANQUILLA	2763	6	1,48 ▲
LDV - BELLO	1533	6	4,70 ▲
LDV - BUCARAMANGA	993	6	1,45 ▲
LDV - CARTAGENA	4834	6	1,64 ▲
LDV - CERETÉ	10179	6	2,23 ▲
LDV - CUCUTÁ	660	6	1,88 ▲
LDV - EL BANCO	171	6	1,33 ▲
LDV - FLORENCIA	2305	6	1,53 ▲
LDV - FUNDACIÓN	2142	6	1,81 ▲
LDV - IBAGÜÉ	2581	6	2,62 ▲
LDV - LA DORADA	115	6	14,39 ▼
LDV - MANIZALES	2578	6	8,78 ▼
LDV - NEIVA	113	6	1,27 ▲
LDV - PASTO	12365	6	1,56 ▲
LDV - POPAYÁN	754	6	2,45 ▲
LDV - SINCELEJO	1445	6	2,69 ▲
LDV - SOGAMOSO	616	6	1,89 ▲
LDV - TULLUÁ	186	6	1,98 ▲
LDV - VALLEDUPAR	34	6	1,06 ▲
LDV - VILLAVICENCIO	2555	6	2,74 ▲
LDV - YOPAL	282	6	1,55 ▲
LNEV	2646	6	12,51 ▼
<b>Total</b>	<b>59223</b>	<b>6</b>	<b>2,84 ▲</b>

Ruta crítica análisis / analistas		
Laboratorio	No. de muestras	Cantidad de días
LDV - AGUACHICA	2230	3,75
LDV - ARAUCA	864	2,03
LDV - ARMENIA	4129	1,81
LDV - BARRANCABERMEJA	149	1,65
LDV - BARRANQUILLA	2763	1,48
LDV - BELLO	1533	4,70
LDV - BUCARAMANGA	993	1,45
LDV - CARTAGENA	4834	1,64
LDV - CERETÉ	10179	2,23
LDV - CUCUTÁ	660	1,88
LDV - EL BANCO	171	1,33
LDV - FLORENCIA	2305	1,53
LDV - FUNDACIÓN	2142	1,81
LDV - IBAGÜÉ	2581	2,62
LDV - LA DORADA	115	14,39
LDV - MANIZALES	2578	8,78
LDV - NEIVA	113	1,27
LDV - PASTO	12365	1,56
LDV - POPAYÁN	754	2,45
LDV - SINCELEJO	1445	2,69
LDV - SOGAMOSO	616	1,89
LDV - TULLUÁ	186	1,98
LDV - VALLEDUPAR	34	1,06
LDV - VILLAVICENCIO	2555	2,74
LDV - YOPAL	282	1,55
<b>Total</b>	<b>59223</b>	<b>3,18</b>

Nota: Elaboración propia

En esta ilustración se comprueba el desempeño desde cuando se recibe la solicitud de las pruebas de laboratorio hasta cuando se entrega los resultados tanto de los laboratorios como de sus analistas indicando si están por encima o debajo de los tiempos establecidos.

## Conclusiones

De acuerdo con la necesidad de análisis del desempeño de laboratorios, muestras y control de la Brucelosis en tiempo real para Colombia, se establece que el número de muestras con resultados positivos es de solo un 3% de las muestras que se han ingresan a SISLAB, el mayor motivo de solicitud de muestras es para el proyecto de certificados de fincas libres de enfermedades de control oficial con un 57,4% de las muestras, esto nos indica que el 3% es el índice promedio que se debe mantener en la línea de tiempo, teniendo en cuenta que el número de casos positivos es directamente proporcional a la cantidad de muestras ingresadas al sistema.

De acuerdo con el proceso realizado de Big data hasta la fecha de corte (15 de abril) se puede identificar que el mayor número de muestras procesadas son provenientes del departamento de Bolívar (13.619), seguido de Nariño (11.290) y Córdoba (6.939) sumando más del 50% del total de las muestras, adicionalmente se pudo identificar que la zona del país con mayor número de animales infectados es el departamento de Nariño con un 48,9% de las muestras positivas.

El procesamiento de las muestras en los diferentes laboratorios se da en mayor volumen en el 1er y 4to trimestre con un 37.6% y 32.8% correspondientemente; los laboratorios con mayor número de muestras analizadas son el laboratorio seccional de Pasto y Cereté con un tiempo de oportunidad de respuesta dentro del tiempo máximo establecido (6 días hábiles).

Se puede concluir que el mayor número de muestras positivas mantiene su tendencia en las hembras, aunque desde el 4to Trimestre del 2021 la mayor cantidad de muestras que se procesaron fueron tomadas en machos. Adicionalmente el mayor número de muestras tomadas se encuentran en un intervalo de edad de 20 a 60 meses y por consiguiente el mayor volumen de muestras con resultados positivos se encuentran en ese mismo rango de edad.

De acuerdo con el tipo de solicitud de las muestras ingresadas en los laboratorios se pudo observar que las muestras particulares mantienen una tendencia estable, mientras que las solicitudes oficiales generan un incremento en el 1er y 4to trimestre.

### **Recomendaciones y trabajos futuros**

- Socializar la estrategia de analítica de datos con el equipo de Big Data del ICA.
- La estrategia de analítica de datos debe ser utilizada para la toma de decisiones de los directores técnicos de la subgerencia de análisis y diagnóstico y protección animal
- Este modelo de arquitectura de datos se puede implementar en las demás enfermedades de control de oficial
- La tabla intermedia creada en SISLAB se debe alimentar cuando las muestras se encuentren liberadas y con un reporte de resultados

## Referencias

Cámara de Comercio de Cali. (2022). Guía básica ¿Qué es la analítica de datos?

<https://www.ccc.org.co/inc/uploads/2019/05/Guia-analitica-de-datos.pdf>

European Knowledge Center for Information Technology. (2015). Microsoft Azure. TIC Portal.

<https://www.ticportal.es/temas/cloud-computing/microsoft-cloud/microsoft-azure>

Fundación ABD. (s.f). Objetivos de desarrollos sostenibles. [https://abd.org/general/abd-recibe-](https://abd.org/general/abd-recibe-la-certificacion-internacional-de-consecucion-de-los-objetivos-de-desarrollo-sostenible-de-las-naciones-unidas/?gclid=EAiaIQobChMIkdHbpJ-_AIV-cqUCR1gfgEgEAAYASAAEgJIR_D_BwE)

[la-certificacion-internacional-de-consecucion-de-los-objetivos-de-desarrollo-sostenible-](https://abd.org/general/abd-recibe-la-certificacion-internacional-de-consecucion-de-los-objetivos-de-desarrollo-sostenible-de-las-naciones-unidas/?gclid=EAiaIQobChMIkdHbpJ-_AIV-cqUCR1gfgEgEAAYASAAEgJIR_D_BwE)

[de-las-naciones-unidas/?gclid=EAiaIQobChMIkdHbpJ-\\_AIV-](https://abd.org/general/abd-recibe-la-certificacion-internacional-de-consecucion-de-los-objetivos-de-desarrollo-sostenible-de-las-naciones-unidas/?gclid=EAiaIQobChMIkdHbpJ-_AIV-cqUCR1gfgEgEAAYASAAEgJIR_D_BwE)

[cqUCR1gfgEgEAAYASAAEgJIR\\_D\\_BwE](https://abd.org/general/abd-recibe-la-certificacion-internacional-de-consecucion-de-los-objetivos-de-desarrollo-sostenible-de-las-naciones-unidas/?gclid=EAiaIQobChMIkdHbpJ-_AIV-cqUCR1gfgEgEAAYASAAEgJIR_D_BwE)

Instituto Colombiano agropecuario –ICA. (2005). Sub. Análisis y diagnóstico.

<https://www.ica.gov.co/areas/subgerencia-analisis-y-diagnostico>

Instituto Colombiano agropecuario- ICA. (2020). Sub. Protección animal.

<https://www.ica.gov.co/areas/pecuaria>)

Instituto Colombiano agropecuario ICA. (s.f.). Laboratorio Nacional de Diagnóstico Veterinario

- LNDV. [https://www.ica.gov.co/areas/subgerencia-analisis-y-diagnostico/laboratorio-](https://www.ica.gov.co/areas/subgerencia-analisis-y-diagnostico/laboratorio-nacional-de-diagnostico-veterinario)

[nacional-de-diagnostico-veterinario](https://www.ica.gov.co/areas/subgerencia-analisis-y-diagnostico/laboratorio-nacional-de-diagnostico-veterinario)

Instituto Colombiano agropecuario ICA. (s.f.). Enfermedades Animales.

<https://www.ica.gov.co/areas/pecuaria/servicios/enfermedades-animales>

Instituto Colombiano agropecuario ICA. (s.f.). Brucelosis Bovina.

[https://www.ica.gov.co/getdoc/1bbc8e4f-12fb-4df0-825a-2f07b8a42367/brucelosis-](https://www.ica.gov.co/getdoc/1bbc8e4f-12fb-4df0-825a-2f07b8a42367/brucelosis-bovina-(1).aspx)

[bovina-\(1\).aspx](https://www.ica.gov.co/getdoc/1bbc8e4f-12fb-4df0-825a-2f07b8a42367/brucelosis-bovina-(1).aspx)

Instituto Colombiano agropecuario ICA. (s.f.). Historia. <https://www.ica.gov.co/el-ica/historia>

- Instituto Colombiano agropecuario ICA. (s.f.). Política de Privacidad y Protección de Datos Personales <https://www.ica.gov.co/home/politica-de-privacidad-y-proteccion-de-datos-perso>
- López Guarnizo, P. (2014). Estudio descriptivo de la presentación de brucelosis humana en Colombia desde 2000 hasta 2012. *Revista de Medicina Veterinaria* (28), 67-79. <http://www.scielo.org.co/pdf/rmv/n28/n28a07.pdf>
- Organización Naciones Unidas. (s.f.). La Declaración Universal de Derechos Humanos. <https://www.un.org/es/about-us/universal-declaration-of-human-rights>
- Power Data. (2021). Big Data y Business Intelligence: el motor de los negocios 3 inteligentes. <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/big-data-y-business-intelligencecrean-negocios-inteligentes-y-rapidos>
- Puyol Moreno, J. (2014). Una aproximación a Big Data. *Revista De Derecho De La UNED (RDUNED)*, (14), 471–506. <https://doi.org/10.5944/rduned.14.2014.13303>
- Rivera, H.G., Benito, A.Z., Ramos, O.C., Manchego, A.S. (2003). Prevalencia de enfermedades de impacto reproductivo en bovinos de la estación experimental de trópico del centro de investigaciones. *Rev Inv Vet Perú*. 15(2): 120–126
- Rochina, P. (2017). Guía para construir un DataWarehouse. Caso práctico. *Revista digital Inesem*, <https://revistadigital.inesem.es/informatica-y-tics/guia-construir-datawarehouse/>
- TIC Portal. (2015). Microsoft Azure. <https://www.ticportal.es/temas/cloud-computing/microsoft-cloud/microsoft-azure>
- Vega Medellín, D. M. (2006). Brucella abortus: antecedentes y avances en aspectos de patogenesis, diagnóstico y control. <https://repository.javeriana.edu.co/bitstream/handle/10554/8252/tesis238.pdf?sequence=1>