



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

AUTOMATIZACIÓN DEL PROCESO DE DATOS E  
IMPLEMENTACIÓN DE DASHBOARDS  
ASOCIADOS A LA PRODUCCIÓN DE PAVOS  
MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL  
INDUSTRIAL

TOMÁS IGNACIO CONTRERAS BASCUÑÁN

PROFESORA GUÍA:  
MARÍA FERNANDA VARGAS COURBIS

MIEMBROS DE LA COMISIÓN:  
GASTÓN SUAREZ CROTHERS  
CAROLINA SEGOVIA RIQUELME

SANTIAGO DE CHILE  
2024

**AUTOMATIZACIÓN DEL  
PROCESO DE DATOS E  
IMPLEMENTACIÓN DE  
DASHBOARDS  
ASOCIADOS A LA  
PRODUCCIÓN DE PAVOS**

**RESUMEN DE LA MEMORIA PARA OPTAR  
AL TÍTULO DE:** Ingeniero Civil Industrial  
**ESTUDIANTE:** Tomás Ignacio Contreras  
Bascañán  
**FECHA:** 2024  
**PROFESORA GUÍA:** María Fernanda Vargas

Agrosuper es una empresa que busca avanzar en la transformación digital mediante la automatización de los procesos de datos relativos a los ciclos productivos de animales, con el fin de fortalecer la toma de decisiones y su capacidad de responder a incidentes mediante la actualización constante de los indicadores de sus dashboards. Uno de sus proyectos radica en la automatización del proceso de datos e implementación de dashboards en la producción de pavos, cuya transición a los datos provenientes de un proceso automatizado desde uno donde la carga de datos era establecida manualmente, tiene el potencial de reproducir todos los paneles existentes para tal ciclo productivo.

El objetivo del proyecto es "*Implementar dashboards asociados a la producción de pavos desde un proceso de datos automatizado, a fin de mejorar la actualización y carga de información para el monitoreo de indicadores por parte del personal de Sopraval*". Para ello, se desarrolla una vinculación de los datos que se cargaban manualmente y los existentes en el proceso automatizado para la obtención de los indicadores prioritarios del ciclo. Posteriormente, los datos actualizados son cargados en un nuevo documento Power BI, en el que se replican los paneles existentes mejorándolos respecto a su versión anterior. Finalmente, se valida el nuevo panel desarrollado y este es publicado para que pueda ser utilizado por los beneficiarios de Sopraval, dando como resultado visualizaciones que se actualizan en promedio veinte veces más cada semana respecto a su versión original y propiciando la mejora continua del desarrollo del proyecto al generar documentación del procedimiento empleado.

A partir de las iteraciones del desarrollo del proyecto, se logra replicar a cabalidad veintiún paneles, mejorando su tasa de actualización y facilitando la toma de decisiones de sus usuarios, además otorgar robustez al proceso al prescindir de la carga manual de datos en los dashboards y aumentar la capacidad de reacción a incidentes en Sopraval. Este proyecto representa un hito significativo para Agrosuper, puesto que su implementación exitosa se traduciría en el primer proceso completamente automatizado en la empresa, considerando desde la ingesta de datos hasta la actualización constante de los paneles, lo que establece un precedente y emerge como una herramienta capaz de generar valor en el resto de las áreas del negocio.

# Agradecimientos

Agradezco a Dios y a todas las personas que me acompañaron en este proceso, a mi familia, amigos que hice en el camino y a todos los que me alentaron a seguir.

Agradezco a la empresa, donde siempre me sentí apoyado y tuvieron la disposición a enseñar e integrarme dentro del equipo.

Agradezco a mis profesores del curso de titulación, quienes siempre mostraron disposición y vocación durante el proceso.

Especialmente dedico este logro a mi tata Carlos Bascuñán, a quien amo y extraño con todo mi corazón y sé que estaría muy orgulloso.

# Tabla de Contenido

1. Introducción .....	1
1.1. Industria y ambiente competitivo .....	1
1.2. Contexto general de la empresa .....	2
2. Desarrollo.....	4
2.1. Justificación del problema y oportunidad de mejora.....	4
2.2. Justificación del proyecto .....	5
2.3. Rol del estudiante .....	6
2.4. Objetivo general.....	7
2.5. Objetivos específicos.....	7
2.6. Alcances .....	8
2.7. Marco Conceptual.....	9
2.8. Metodología .....	13
2.9. Desarrollo del proyecto.....	19
2.10. Resultados .....	30
3. Conclusiones.....	34
Bibliografía .....	37
Anexo A .....	38
A.1. Certificaciones segmento Carnes .....	38

# Índice de Ilustraciones

Ilustración 1: Ventas totales por negocio.....	2
Ilustración 2: Proceso de datos .....	9
Ilustración 3: Modelo Snowflake.....	11
Ilustración 4: Proceso de datos Agrosuper .....	13
Ilustración 5: Ciclo productivo de pavos.....	14
Ilustración 6: Tipos de tablas de hechos.....	15
Ilustración 7: Proceso metodológico.....	16
Ilustración 8: Tabla comparativa nacimiento.....	22
Ilustración 9: Ejemplo estructura dimensional tabla Mortalidad.....	24
Ilustración 10: Ejemplo panel de temática Mortalidad.....	26
Ilustración 11: Ejemplo panel temática Cierres.....	27
Ilustración 12: Paneles abordados por temática.....	30
Ilustración 13: Visualización nacimiento a replicar.....	31
Ilustración 14: Visualización nacimiento replicada.....	32

# 1.Introducción

## 1.1. Industria y ambiente competitivo

Agrosuper SPA. es una empresa chilena que se dedica a la producción y comercialización de alimentos, especialmente en las industrias de carne de ave (avícola), porcina y acuícola. La empresa opera bajo diferentes marcas y líneas de productos en el mercado, donde algunas de las más conocidas son Super Pollo, Super Cerdo, SopraVal, La Crianza y AquaChile.

A nivel nacional, la empresa está inmersa en tres negocios principales (Agrosuper, 2023, p. 70):

1. Carne de pollo: donde su participación de mercado es de un 50% considerando la importación y producción de dicha carne.
2. Carne de cerdo, donde la participación de mercado de la empresa es de un 54% considerando su importación y producción.
3. Carne de pavo: donde su participación de mercado es de un 65% bajo las mismas condiciones.

A nivel internacional, Agrosuper influye en el mercado global a través de cuatro líneas de negocio principales (Agrosuper, 2023, p. 71):

1. Salmón Atlántico: en este segmento, AquaChile representa el 5,2% del mercado global de salmón Atlántico.
2. Salmón del Pacífico: AquaChile se posiciona con un 24,2% de cuota en el mercado mundial de salmón del Pacífico.
3. Carne de ave: en el mercado mundial de aves, considerando la carne de pollo y pavo, Agrosuper aporta un total del 0,4%.
4. Carne de cerdo: Agrosuper cuenta con una participación del 0,3% en el mercado internacional de carne de cerdo.

## 1.2. Contexto general de la empresa

La Sociedad Agrosuper SPA. es controlada por Gonzalo Vial Vial, y por sus cuatro hijos, los que mediante las sociedades Agrocomercial El Paso S.A., Promotora Doñihue LTDA., Agrícola GV S.A. e Inversiones VC LTDA. son titulares del 100% de las acciones de la empresa de manera directa e indirecta.

Desde su fundación en el año 1955, la empresa se ha convertido en un actor principal en la industria alimentaria a nivel nacional e internacional. Actualmente, la empresa cuenta con veintinueve sucursales y tiendas repartidas a nivel nacional y con once oficinas comerciales de nivel internacional repartidas en Chile, Estados Unidos, Italia, China, Corea del Sur y Japón. Lo que se traduce en que los productos de la empresa se comercialicen en 64 países en el presente.

Según el reporte económico y financiero más reciente, correspondiente al año 2022, la empresa registró ventas totales superiores a 4.179 millones de dólares, de las que el 39,7% son ventas nacionales, las que proporcionan un monto cercano a 1.658 MUS\$ y el 60,3% son internacionales, cuyo valor asciende a 2.521 MUS\$ aproximadamente (Agrosuper, 2023, p. 52).

Adicional a lo anterior, las ventas totales se distribuyen por tipos de negocio según lo siguiente:

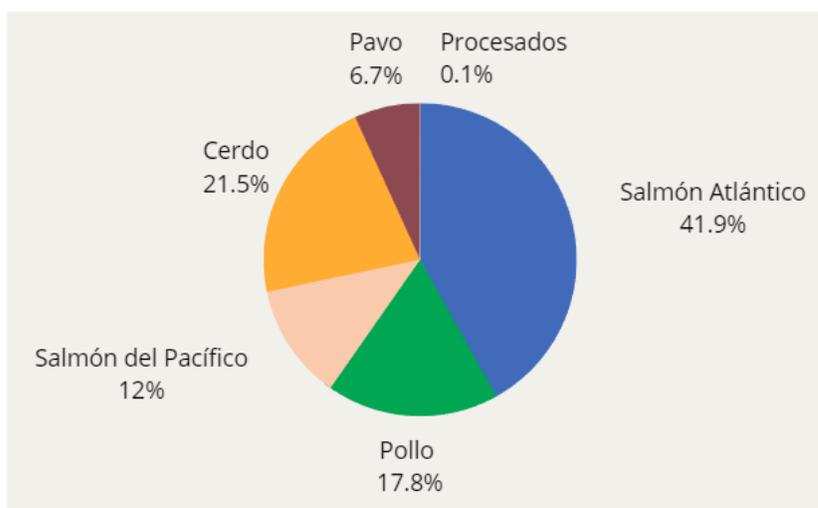


Ilustración 1: Ventas totales por negocio (Fuente: Elaboración propia).

Una particularidad de Agrosuper es que su cadena de valor se encuentra verticalmente integrada, es decir, controla diversas etapas del proceso de producción, desde la cría de los animales hasta la distribución de los productos terminados. Esta integración le permite a Agrosuper mantener control en la

calidad y eficiencia de sus operaciones y contar con certificaciones de estándar internacional. (Anexo A.1.)

Dentro de la estructura de la empresa, se destacan varias áreas, cada una con un papel esencial en el funcionamiento integral de la compañía. Estas áreas incluyen Marketing, Administración y Finanzas, Industrial, Innovación, Personas, Asuntos Corporativos, Ventas y Producción Animal. La relevancia de la última área, Producción Animal, es particularmente notable. Esta se dedica a la crianza y reproducción de aves y cerdos, cuyo destino final es el consumo humano. En este contexto, se consideran aspectos como el bienestar de los animales, la bioseguridad y el desarrollo del personal implicado en esta área.

Cada una de las áreas previamente mencionadas tiene un departamento de transformación digital, el que cuenta con un equipo de Analítica y BI, cuya principal función es validar la consistencia de los datos y generar herramientas que faciliten la interpretación y el acceso a la información a partir de estos. En el área de Producción Animal, dichas herramientas muestran indicadores relacionados con las diferentes etapas de vida de los animales, cuyos datos fueron previamente recopilados directamente desde las granjas, pabellones y plantas correspondientes. Por ende, el detectar incidentes a tiempo y poder reaccionar oportunamente a estos es fundamental para el funcionamiento de Agrosuper.

Existen indicadores prioritarios y correspondientes a cada etapa del proceso productivo de los animales, los que deben ser abordados según los requerimientos de los encargados de las etapas del ciclo productivo para cada animal y mejorados continuamente en base a las nuevas necesidades o perspectivas que existan para monitorear un indicador. Para cada proceso productivo de los que se encarga el área de Producción Animal, el equipo de Analítica y BI debe mantener los datos actualizados para generar dashboards que faciliten la interpretación y monitoreo de los indicadores correspondientes.

## 2. Desarrollo

### 2.1. Justificación del problema y oportunidad de mejora

Actualmente, el departamento de transformación digital del área de Producción Animal presenta el desafío de automatizar el proceso de obtención de los datos que alimentan las visualizaciones de los indicadores, particularmente en el ámbito de la producción de pavos por parte de Sopraval. El objetivo principal es mantener los datos constantemente actualizados, mejorando el monitoreo de los indicadores y agilizando la capacidad de responder a incidentes. Adicionalmente, se busca prescindir de la carga manual de los datos por un miembro del equipo y evitar posibles errores vinculados a la carga manual de estos en las etapas del ciclo productivo.

En la situación actual, se cargan los datos actualizados para generar las visualizaciones de los indicadores de forma semanal y manual, los que son obtenidos desde un Excel. Esto se traduce en dificultades para identificar a tiempo las problemáticas que puedan existir en las plantas, granjas y pabellones, generando respuestas tardías a situaciones críticas para el negocio, específicamente en los indicadores prioritarios del proceso productivo. Además, esta actualización de los datos requiere necesariamente la intervención de los integrantes del equipo de Analítica y BI, quienes dedican parte de su tiempo cada semana solo para mantener los indicadores existentes en las visualizaciones actualizados.

Sumado a lo anterior, el área de Producción Animal se encuentra en una fase de reajuste en la dotación de personal, lo que ha generado traslados de los integrantes del equipo a otros puestos de trabajo, incluyendo al Ingeniero que se dedicaba en realizar el proceso de la carga de los datos del ciclo productivo de los pavos, quien se encarga de desempeñar las funciones de su nuevo puesto de trabajo como algunas de las tareas de su cargo anterior.

Los factores mencionados intensifican la urgencia de abordar el desafío de transformación digital, ya que la necesidad de supervisar el proceso de producción persiste y la oportunidad de automatizar el proceso mediante el que se accede a los datos permite abordar estas problemáticas de manera transversal, mejorando la constancia con la que se actualizan los indicadores, prescindiendo de la carga manual de los datos y respondiendo a su vez a las dificultades generadas por el proceso de reajuste.

## 2.2. Justificación del proyecto

El proyecto en el que el estudiante se involucra se centra en la “Automatización del proceso de carga de datos y desarrollo de Dashboards relacionados con la producción de pavos”. El responsable de este proyecto es el Jefe de Analítica y BI, y cuenta con la participación de todo el equipo de Analítica y BI dentro del área de Producción Animal.

Este proyecto se enfoca en utilizar las bases de datos alojadas en la plataforma Microsoft Azure. A través de Azure Synapse Analytics, estas bases de datos se segmentan para ser conectadas con los paneles de Power BI, los cuales a su vez alimentan los dashboards que exhiben los indicadores asociados a la producción de pavos. Aunque los datos obtenidos a través de esta metodología son esencialmente los mismos que se obtendrían mediante la carga manual desde Excel, la implementación de este proyecto ofrece una ventaja significativa. Al establecer la conexión entre Azure Synapse y Power BI, los dashboards se actualizan automáticamente cada vez que se ingresan nuevos datos en Microsoft Azure.

Esta automatización elimina la necesidad de la labor manual que realiza actualmente el ingeniero encargado de los pavos. Además, posibilita la generación constante y permanente de visualizaciones para cada indicador, abor-dando por completo la problemática mencionada anteriormente. Esta implementación también contribuye al desarrollo de un sistema de monitoreo más efectivo.

Es importante mencionar que existen proyectos similares para la producción de pollos y cerdos, los cuales están parcialmente implementados. Adicionalmente, los departamentos de transformación digital en diversas áreas de Agro-super tienen como objetivo seguir avanzando en esta dirección de automatización en el futuro.

La realización de este proyecto representa un hito significativo tanto para la empresa como para el equipo involucrado. Si esta metodología se implementa con éxito, se lograría el primer proceso completamente automatizado dentro de la empresa, considerando desde la carga inicial de datos hasta la actualización constante de los dashboards.

## 2.3. Rol del estudiante

La participación del estudiante, quien se integra al equipo con el rol de un Ingeniero en Analítica y BI, considera labores en tres etapas principales del proyecto, cada una de ellas fundamental para su realización exitosa. Además, se deben considerar aspectos transversales que se detallan a continuación.

En la primera etapa, la labor del alumno consiste en establecer un mapeo de información entre los datos actualmente utilizados en las visualizaciones y los datos almacenados en las tablas de Microsoft Azure Synapse, que se detallarán más adelante. Para lograr esto, es crucial contar con un profundo entendimiento del proceso de producción de pavos, así como manejar las metodologías para calcular los indicadores presentes en las vistas de Power BI. Esto permitirá identificar los datos necesarios para generar visualizaciones relevantes, descartando aquellos que no aporten información.

En la segunda etapa, el estudiante debe automatizar el proceso de acceso a los datos, asegurándose de validar la coherencia entre las bases de datos existentes y posteriormente integrándolos en Power BI. Además, será responsabilidad del estudiante establecer un modelo lógico para utilizar las tablas, generando relaciones funcionales entre ellas.

La tercera etapa implica la reproducción de las visualizaciones ya existentes utilizando las tablas que previamente se cargaron desde Azure Synapse. Esto requerirá un sólido dominio de Power BI por parte del estudiante y la capacidad de utilizar datos cargados y dispuestos de manera diferente a la de los dashboards a replicar.

De manera transversal a las etapas mencionadas, el alumno debe aportar al análisis de la eficacia y eficiencia de la metodología empleada durante el desarrollo del proyecto. Además, se espera que el estudiante proponga visualizaciones más efectivas para interpretar la información de los paneles y documente las acciones realizadas a lo largo del proceso. En todo momento, será esencial una comunicación fluida con los miembros del equipo de Sopraval para entender y atender sus requerimientos de manera oportuna.

## 2.4. Objetivo general

Implementar dashboards asociados a la producción de pavos en el área de Producción Animal desde un proceso de datos automatizado, a fin de mejorar la actualización y carga de información para el monitoreo de indicadores por parte del personal de Sopraval.

## 2.5. Objetivos específicos

- Crear una tabla comparativa que vincule las fuentes de origen de información, estableciendo un mapeo de los datos para el posterior cálculo de indicadores a partir de la nueva fuente de origen.
- Generar un nuevo documento Power BI, con datos provenientes de la nueva fuente de origen correspondiente al proceso de datos automatizado y validando que los datos cuadren con la fuente de origen anterior.
- Desarrollar los nuevos paneles relativos al proceso de producción de pavos, estableciendo mejoras en las visualizaciones y en las tasas de actualización respecto a la versión anterior.
- Documentar el proceso de desarrollo del proyecto, analizando la eficiencia y efectividad de las metodologías empleadas y propiciando la mejora continua del proyecto.

## 2.6. Alcances

El proyecto mencionado abarca desde la obtención de los datos disponibles en las tablas de Azure Synapse hasta que las nuevas visualizaciones son generadas en Power BI a partir de dichos datos. Lo anterior, validando que los datos presentes en las tablas se estén ingresando de manera adecuada y reportando a los encargados de almacenar los datos en caso de encontrar diferencias respecto a los datos de referencia.

Las nuevas visualizaciones generadas son entregadas al personal de Sopraval, quienes son los encargados de monitorear los indicadores respectivos a cada parte del proceso y reaccionar a los incidentes que se puedan presentar en las plantas, granjas y pabellones. Las tablas de Azure Synapse por su parte, contienen toda la información necesaria para replicar los paneles existentes en base a la nueva fuente de información, correspondientes a todas las etapas del ciclo productivo de pavos que será esquematizado más adelante.

Considerando que los datos incorporados en las tablas se están almacenando recientemente, es recurrente que algunos de los datos necesarios para generar los indicadores no se estén ingresando correctamente, lo que impide la reproducción precisa de algunas visualizaciones. Por consiguiente, el progreso del proyecto y la participación del estudiante en el mismo dependen directamente de la meticulosidad y celeridad con que se incorporen los datos en la base.

Pese a lo descrito en el párrafo anterior, posterior a que se corrijan las deficiencias en cómo se ingresan los datos y las diferencias existentes en los datos, la información contenida en estos constituye todo lo necesario para replicar con exactitud los indicadores presentes en los paneles originales. Por lo tanto, el alcance del proyecto radica en generar todas las vistas solicitadas por la contraparte durante la estadía del estudiante en la empresa exclusivamente desde los datos pertenecientes a las tablas de Azure Synapse, en otras palabras, con un 100% de éxito.

El alumno en práctica estará presente desde el momento que comienza el proyecto, por lo que las peticiones de visualizaciones a generar por parte de Sopraval corresponden a las que aportan mayor valor en cada etapa del proceso de producción. Debido a que la extensión de tiempo que tomará el proyecto es incierta y considerando que las visualizaciones generadas estarán sujetas a una mejora continua, es necesario que el alumno documente el desarrollo a lo largo del proceso de práctica con el fin de facilitar la continuidad del proyecto posterior a la finalización de su estadía en la empresa.

## 2.7. Marco Conceptual

El proyecto se encuentra inmerso en las disciplinas de Análisis de datos e Inteligencia de Negocios, el eje central está en rediseñar los paneles de datos, migrando las fuentes de datos manuales a fuentes que son almacenadas de forma automática en el Data Lake, lo que implica la obtención de un proceso de datos automatizado. En la presente sección, se profundizará respecto a los modelos aplicados a lo largo del proyecto y su marco teórico correspondiente.

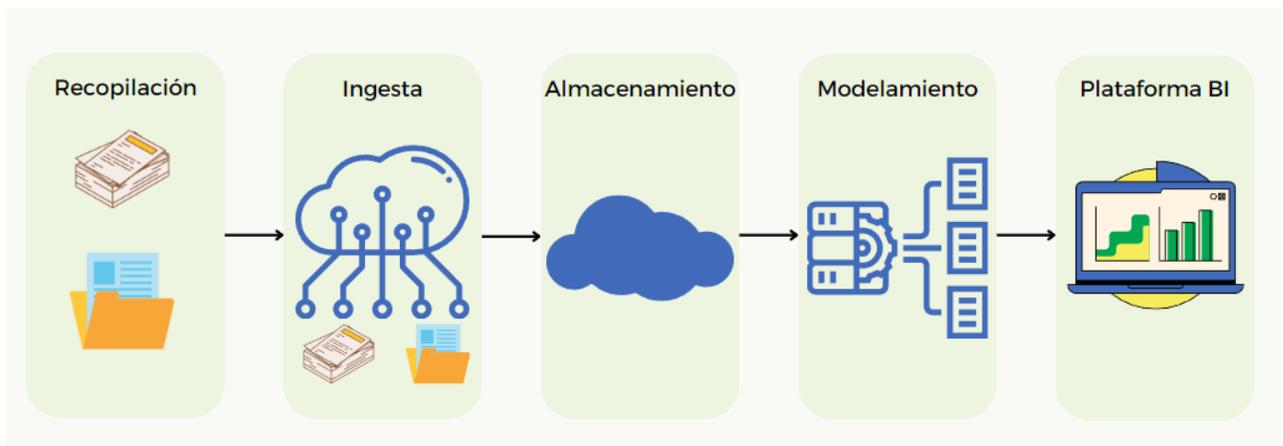
### 2.7.1 Proceso de datos

Para entender donde se sitúa el proyecto en el que trabajará el estudiante, es necesario entender el funcionamiento del proceso de datos utilizados para generar los indicadores referentes al proceso productivo. A continuación, se explicarán los principales conceptos ligados a dicho proceso.

#### 1. Data Pipeline

El data pipeline corresponde la cadena compleja de actividades o procesos interconectados desde que se generan los datos hasta que se reciben, donde el resultado de uno o más procesos se convierte en un insumo para otro (Raman K. *et al*, 2013). Para este propósito es necesario orquestar las etapas del proceso para que estén coordinadas en la automatización. Esto permite que el flujo de datos ingestados en tiempo real, al mismo tiempo, alimente el punto de llegada para aportar al monitoreo y la toma de decisiones.

En la siguiente imagen, se muestra un proceso de datos esquematizado desde la recopilación hasta la carga de datos en una plataforma BI:



*Ilustración 2: Proceso de datos (Fuente: Elaboración propia).*

De manera general, los datos recopilados al comienzo del proceso son ingeridos para luego ser almacenados en la nube, en donde se incorporan los datos de manera directa y masiva, sin ser procesados. Posterior a su almacenamiento, es necesario modelar estos datos y procesarlos para que puedan ser cargados en las plataformas BI, donde son utilizados para generar paneles que permiten visibilizar los indicadores necesarios para un negocio.

## 2. Data Lake

El concepto de Data Lake se refiere a un repositorio de almacenamiento escalable y masivo de datos que contiene una gran cantidad de datos sin procesar, en un formato nativo hasta que se necesiten herramientas para procesarlos, de modo que se puedan ingerir sin compromiso de su estructura (Milovskaya & Tosltoy, 2016).

## 3. Modelamiento de datos

Previo a que los datos provenientes del Data Lake alimenten el punto de llegada, es necesario modelar los datos para establecer una estructura, lo que hace más eficiente el proceso de obtención de los datos y evita ralentizaciones en la plataforma de llegada.

Dentro de las formas de modelar los datos, existen los modelos dimensionales, que corresponden a una manera de mejorar el rendimiento de las consultas resumiendo la información sin afectar la integridad de los datos. En estos, se pueden identificar dos tipos de tablas, las tablas de dimensiones y las tablas de hechos. Las tablas de dimensión contienen los atributos que describen o especifican como deberían resumirse los datos pertenecientes a las tablas de hechos para otorgar información útil. Por su parte, las tablas de hechos contienen los datos que se busca medir o analizar (Ballard C. *et al*, 2016).

El modelo Snowflake corresponde a un tipo de modelo dimensional compuesto por una tabla de hechos (tabla azul) conectada a tablas de dimensión (tablas rojas), además cada tabla de dimensión se conecta a otras tablas dimensionales formando un esquema copo de nieve.

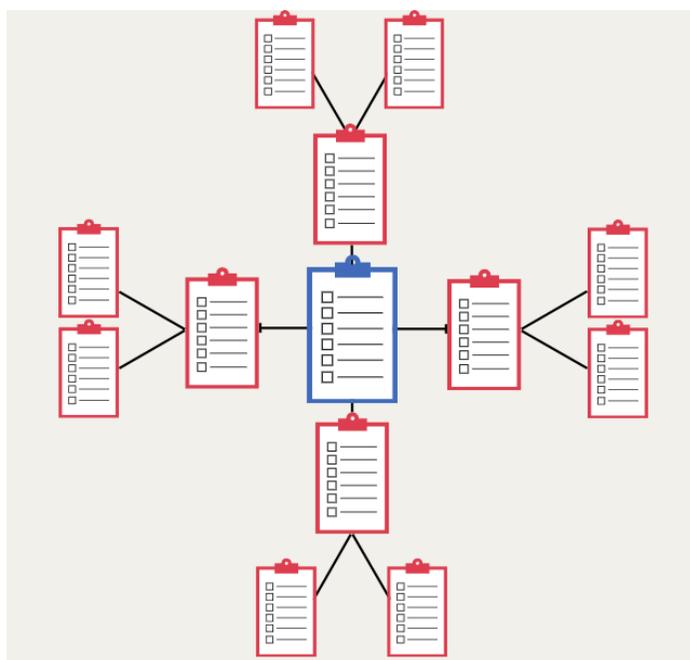


Ilustración 3: Modelo Snowflake (Fuente: Elaboración propia).

## 2.7.2 Inteligencia de Negocios (BI)

La Inteligencia de Negocios (Business Intelligence) se refiere a técnicas informáticas utilizadas para localizar, extraer y analizar datos del negocio, como ingresos por ventas, por productos y/o departamentos, o por costos e ingresos asociados. Las tecnologías de BI proporcionan vistas históricas, actuales y predictivas de las operaciones de un negocio (Elena C. 2011) Luego de introducir este concepto, se abordará sobre los conceptos de performance dashboards y visualizaciones de datos.

### 4. Performance dashboards

Un performance dashboard (panel de desempeño) es un sistema de gestión del desempeño, que comunica objetivos estratégicos y permite al personal de las empresas medir, monitorear y gestionar las actividades y procesos clave para lograr sus objetivos (Eckerson W. 2010). Estos permiten apoyar la toma de decisiones respecto a lo que se esté midiendo, mediante el monitoreo de los indicadores que presente un panel, por lo que, si estos paneles son actualizados con mayor frecuencia, se traduce en facilitar la toma de decisiones por parte del personal y reaccionar rápidamente a los procesos o actividades involucradas.

## 5. Visualizaciones de datos

Dentro de la Inteligencia de Negocios, la visualización de datos es el proceso de representar datos de forma gráfica de manera clara y efectiva (Sadiku M. *et al* 2016). A partir de estas representaciones de datos, es posible visibilizar los datos contenidos en las tablas de hechos, desde diferentes perspectivas (dimensiones) que aporta el modelamiento y monitoreo de indicadores del negocio en los diferentes objetos visuales disponibles.

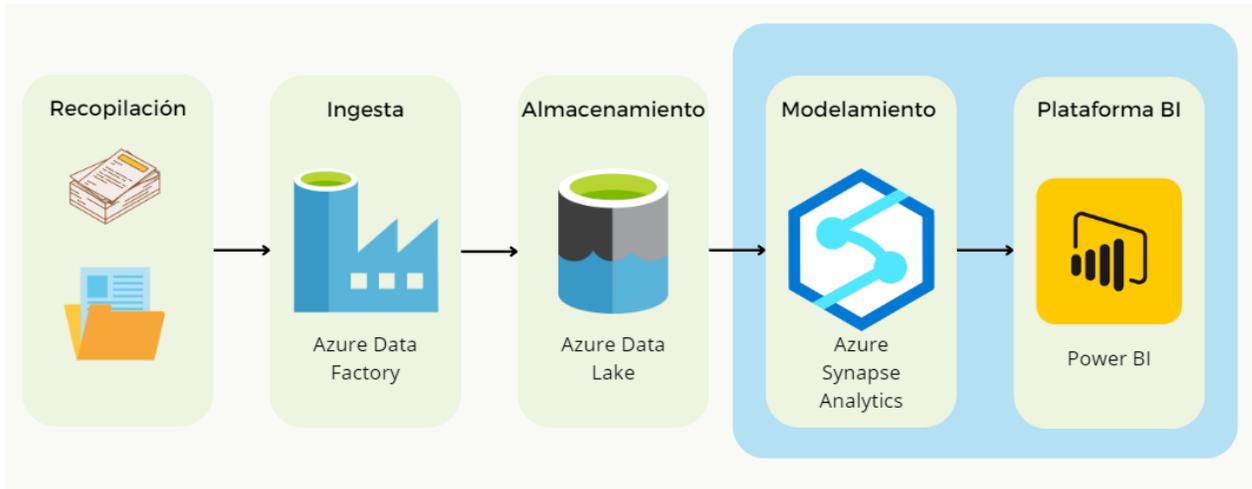
## 6. Key Performance Indicators (KPI)

Los KPI representan un conjunto de medidas que se centran en aquellos aspectos del desempeño organizacional que son los más críticos para el éxito actual y futuro de la organización (Domínguez E. *et al* 2019).

## 2.8. Metodología

En esta sección se explicitan las herramientas utilizadas para llevar a cabo el proyecto, describiendo los motivos por las que fueron seleccionadas y los recursos utilizados en las etapas del desarrollo de este.

A partir de las normativas y buenas prácticas propuestas por Agrosuper, se propone siguiente proceso para la carga de datos en Power BI:

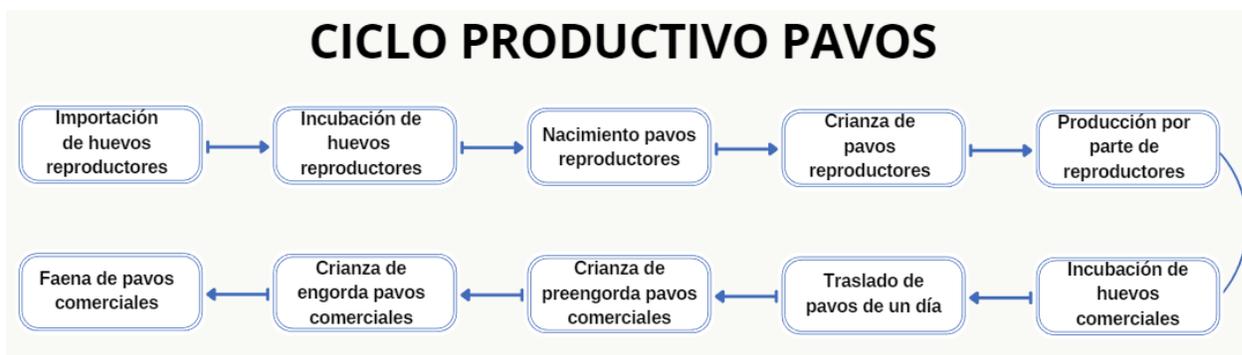


*Ilustración 4: Proceso de datos Agrosuper (Fuente: Elaboración propia).*

Posterior a las etapas de recopilación e ingesta de datos, se tiene que las herramientas utilizadas para el almacenamiento y modelamiento corresponden a Azure Data Lake Storage y Azure Synapse Analytics respectivamente. En cuanto a la plataforma BI utilizada, en todas las áreas de Agrosuper se utiliza Microsoft Power BI, perteneciente a la infraestructura de Microsoft en conjunto con el resto de las herramientas mencionadas. El cuadro celeste, engloba la parte del proceso de datos en donde se encuentra inmerso el proyecto a desarrollar por parte del estudiante y el equipo de Analítica y BI.

### 2.8.1 Recursos e insumos del proyecto

Previo a la descripción del procedimiento empleado en la solución que amerita el proyecto, es necesario contextualizar sobre el ciclo productivo de pavos y hacer unos alcances sobre las tablas de Azure Synapse Analytics aplicadas a dicho proceso.



*Ilustración 5: Ciclo productivo de pavos (Fuente: Elaboración propia).*

En el proceso de producción de pavos es importante notar que el ciclo completo considera dos generaciones, los pavos reproductores y los pavos comerciales. La procedencia de los huevos difiere dependiendo de la generación, por un lado, los reproductores tienen su origen a partir de huevos importados, donde posterior al periodo de incubación, nacen y llegan a la etapa de crianza, para luego producir los huevos desde los que nacerán los pavos comerciales de la siguiente generación.

Luego, tras el nacimiento de los pavos comerciales, estos son trasladados durante su primer día de vida para experimentar dos etapas de crianza, las que corresponden a preengorda (las primeras seis semanas de vida) y engorda (hasta las veinte semanas de vida). Luego de cumplir con dichas etapas, las aves son trasladadas a las plantas faenadoras y finalmente los productos generados son distribuidos para su posterior consumo.

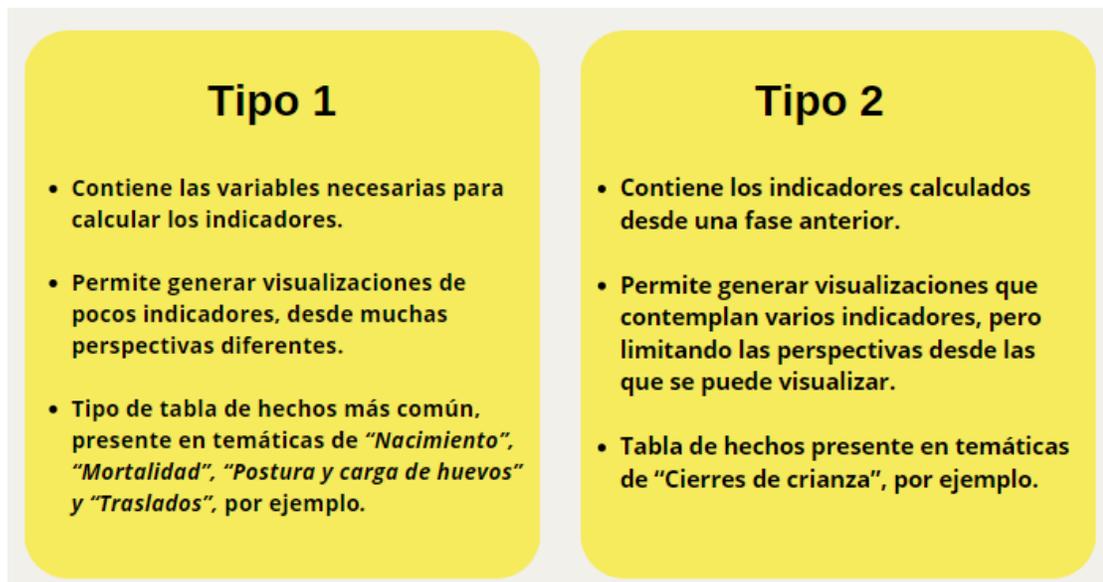
En las diferentes etapas de este ciclo productivo, se tiene una tabla de hechos asociada a cada temática abordada, la que generalmente contiene la información necesaria para obtener todos los indicadores asociados a dicha temática. Las temáticas más recurrentes a lo largo del ciclo productivo: "Postura e incubabilidad", "Mortalidad", "Nacimiento", "Cierres de crianza" y "Traslados".

De manera general, en la situación previa a la realización del proyecto, cada una de estas temáticas tenía asociado un documento Excel, que contiene los datos necesarios para obtener los indicadores asociados a cada una de ellas. Por ejemplo, el documento Excel "*BD Nacimiento*" es el documento que contiene los datos necesarios para obtener indicadores como "*Total de nacidos*" o "*Porcentaje de nacimiento*". Este documento debe ser actualizado periódicamente de manera manual para mantener las visualizaciones originales con los datos más recientes y así los indicadores actualizados.

Análogamente, existe una tabla de hechos ligada a cada una de estas temáticas en Azure Synapse, que contiene la información necesaria para obtener

cada uno de los indicadores respectivos a una temática. Siguiendo con el ejemplo anterior, en Azure Synapse existe la tabla "NACIMIENTO", que aporta con la información necesaria para obtener los indicadores mencionados en el párrafo anterior y permite mantener las visualizaciones actualizadas mediante el proceso de datos automatizado.

En este proceso de datos de la empresa (Ilustración 4), se pueden identificar dos tipos de tablas de hechos según el siguiente esquema:



*Ilustración 6.: Tipos de tablas de hechos (Fuente: Elaboración propia).*

Por un lado, las tablas de hechos del primer tipo corresponden a las que contienen todas las variables requeridas para obtener los indicadores, es decir, será necesario calcular posteriormente los indicadores a partir de dichas variables para poder replicar un panel en la plataforma BI. Por otro lado, las del segundo tipo contienen los indicadores calculados desde una fase anterior del proceso de datos, por lo que en general contienen muchos indicadores diferentes y los objetos visuales a generar, a su vez, muestran muchos indicadores en su conjunto. La distinción entre estas tablas es relevante para el desarrollo del proyecto, como se explicará en el desarrollo del proyecto.

## 2.8.2 Proceso metodológico

A continuación, se describe el proceso metodológico empleado para abordar el proyecto representado mediante el siguiente esquema. Los primeros cuatro pasos de esta metodología (en el cuadro celeste) fueron ejecutados exclusivamente por parte del estudiante.

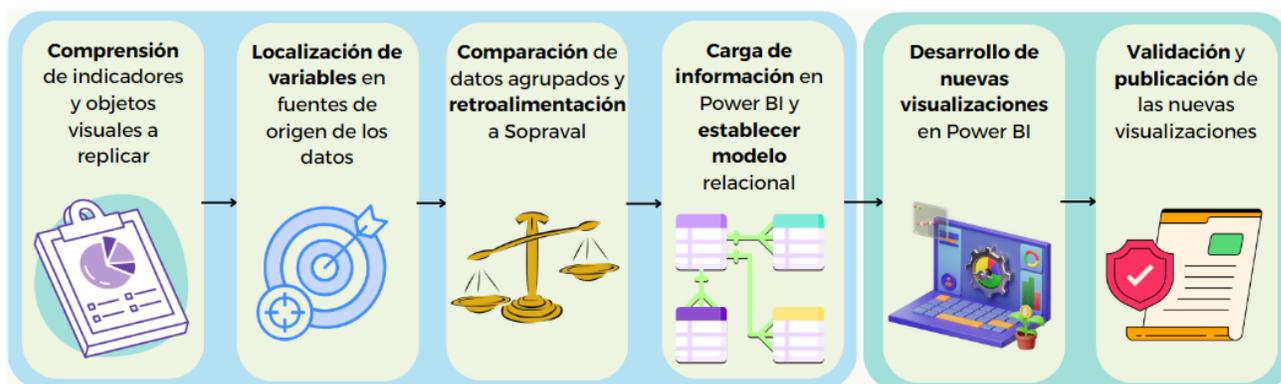


Ilustración 7.: Proceso metodológico (Fuente: Elaboración propia).

### 1. Comprender los indicadores y visualizaciones objetivo a replicar

En esta etapa, la primera tarea a realizar radica en comprender qué visualizaciones de la temática se deben replicar y cuáles son los indicadores necesarios para obtener los objetos visuales, además de identificar las variables que se utilizaron para calcularlos y las diferentes dimensiones o perspectivas que utilizan para monitorear la información. Esta tarea se debe realizar en Power BI, pues en este programa están las visualizaciones a abordar con los indicadores y cálculos pertinentes al ciclo productivo de pavos. Microsoft Power BI es la plataforma BI acordada a utilizarse en todas las áreas de Agrosuper y corresponde a una herramienta que se alimenta de los datos del negocio, permitiendo limpiarlos y transformarlos para generar objetos visuales como gráficos, tablas o paneles que faciliten la interpretación de la información por parte de los usuarios.

### 2. Localizar las variables necesarias en ambas fuentes de origen

Para esta parte del proceso, en primer lugar, es necesario verificar la fuente de origen de datos desde la que se generan los distintos objetos visuales en Power BI, la que corresponde a un documento Excel. Una vez se accede al documento, se debe localizar en donde se encuentran alojadas las variables necesarias para el cálculo de los indicadores (KPI's) e interpretar la lógica con la que se calculan los valores necesarios para generar las visualizaciones.

En segundo lugar, se debe realizar la acción análoga entendiendo las tablas de Azure Synapse, correspondiente a la nueva fuente de origen de datos, que en la mayoría de los casos contienen las variables en una sola tabla de hechos, pero en el caso de no ser así, es necesario lograr obtener las variables necesarias para el cálculo de los indicadores apoyándose de otras tablas. La tarea de consultar las tablas de Azure Synapse se realiza en la herramienta SQL Server Management Studio, que corresponde a un entorno que permite administrar, configurar o consultar las instancias de un servidor o una base de datos. El estudiante utiliza esta herramienta debido a que permite consultar una gran cantidad de tablas diferentes en paralelo y acceder a los datos mediante agrupaciones o aplicando filtros, lo que facilita el proceso de búsqueda.

### 3. Comparar los datos agrupados y retroalimentar a personal de Sopraval

Una vez identificadas las variables en ambas fuentes de origen y agrupando los datos para que sean comparables, estos se incorporan a una hoja de Excel. Así, se forma una tabla comparativa, donde para cada grupo de datos los valores respectivos a una variable deberían ser iguales, pues hacen referencia a datos que significan lo mismo. La importancia de esta comparación radica en identificar en qué grupos de datos existen diferencias y reportar a los encargados de recopilar, ingestar y almacenar la información que parte de los datos no están cargándose adecuadamente en la nueva fuente de origen.

Esta tabla es compartida al personal de Sopraval, explicando las diferencias que existen entre grupos de datos y asegurando que la metodología desde la que se localizaron estos es la correcta. En la sección del desarrollo del proyecto, se mostrará una tabla comparativa como ejemplo.

### 4. Cargar tablas de Synapse en Power BI y establecer el modelo dimensional

En esta etapa, se procede a cargar la información al nuevo documento Power BI que tendrá las visualizaciones replicadas en base los datos provenientes solo desde las tablas de Azure Synapse.

Al estar cargadas todas las tablas necesarias para generar las visualizaciones, se establecen las relaciones funcionales entre las tablas siguiendo el modelo Snowflake. En otras palabras, se conectan las tablas de hechos y las tablas dimensionales de modo que, al momento de generar las visualizaciones, estas se puedan apreciar desde las perspectivas que se necesiten correctamente.

Una vez abarcadas estas etapas, se establece un proceso de datos automatizado, donde las nuevas visualizaciones dependerán solo de los datos cargados en la nueva fuente de origen y la actualización de dichos datos en adelante de acuerdo con el concepto de Data Pipeline expuesto en el marco conceptual. Desde este punto de la metodología, tanto el estudiante como el resto del equipo se encargan de las tareas de las siguientes etapas (marcadas en verde), por lo que cabe mencionar que las principales responsabilidades del alumno están inmersas en el proceso de datos más que en generar las nuevas visualizaciones.

#### 5. Desarrollar las nuevas visualizaciones en Power BI

Tras conectar las tablas, se procede a replicar cada una de las visualizaciones solicitadas, donde se deben imitar rigurosamente las metodologías de cálculo utilizadas en la visualización original y ajustarlas a los nuevos datos en caso de ser necesario. Adicionalmente, una vez generados los nuevos objetos visuales, se aplican segmentaciones y filtros a los datos para que la perspectiva desde la que se monitorea la información sea la requerida por los usuarios de los paneles.

Una vez replicadas las visualizaciones, la tabla comparativa generada anteriormente es utilizada para verificar que, para los grupos de datos que cuadraron entre ambas fuentes de origen, las visualizaciones deben ser idénticas en el Power BI original y el replicado. Esta instancia es fundamental, pues permite verificar que desde que se establecieron las relaciones entre las tablas en adelante se hayan realizado las tareas correctamente

#### 6. Validar y publicar las nuevas visualizaciones

Tras finalizar el proceso de desarrollo de un nuevo documento Power BI, este es validado por el jefe de Analítica y BI, quien solicita realizar los últimos ajustes a las nuevas visualizaciones (generalmente estéticos) y posterior a esto, publica el documento para que los trabajadores de SopraVal tengan acceso. Dependiendo de la visualización, se define la frecuencia con la que se actualizará el nuevo documento publicado y se verifica que los responsables de recopilar, ingestar y almacenar los datos se estén encargando de arreglar las diferencias reportadas en la etapa tres.

Al estar publicadas las nuevas visualizaciones, se solicita reunirse nuevamente con personal relacionado a alguna temática del proceso de producción de pavos, para que se generen los siguientes requerimientos para seguir con el proyecto, repitiéndose el ciclo y las etapas recién expuestas.

## 2.9. Desarrollo del proyecto

En la presente sección, se explicará el desarrollo del proyecto aplicado a un ejemplo de cada tipo de tabla de hechos, debido a que la complejidad de este se concentra en etapas distintas de la metodología dependiendo del tipo de tabla de hechos que se esté utilizando para generar la visualización, como se mostrará próximamente. Además, las secciones del desarrollo hacen referencia a los objetivos específicos planteados, de modo que se explicita en qué parte del desarrollo se cumplen dichos objetivos acorde a la metodología empleada.

En el proceso de desarrollo, se abordaron cinco de las temáticas más recurrentes del ciclo productivo de pavos. Estas contienen los indicadores prioritarios en las solicitudes del personal de Sopraval y en base a estas, se mostrarán los resultados obtenidos a lo largo de la estadía del estudiante en la empresa.

### 2.9.1 Comprensión de indicadores y objetos visuales a replicar

Posterior a que el personal de Sopraval indica al equipo de Analítica y BI del área de Producción Animal cuáles son las visualizaciones por replicar según sus requerimientos prioritarios, se abarca la primera etapa de la metodología.

En un comienzo, se identifica cuáles indicadores son los necesarios para reproducir las visualizaciones pedidas, las variables requeridas para calcularlos y todos los elementos adicionales a considerar para generar una réplica cabal de la visualización a partir de la nueva fuente de origen de información. La ejecución de estas tareas es realizada en una vista del Power BI original, donde se puede acceder a las tablas de datos utilizadas para la creación de las visualizaciones y las que contienen a su vez, todos los cálculos necesarios para generar los indicadores.

Desde las tablas de datos, es posible acceder al nombre del documento Excel exacto desde el que fueron cargados los datos a cada tabla. Lo anterior, permite asegurar desde qué documento se cargó la información para obtener las visualizaciones del documento BI a replicar y comprender de qué forma están dispuestos los datos que dieron origen a los objetos visuales a replicar.

## 2.9.2 Localización de variables en ambas fuentes de origen de datos

Luego de acceder al archivo Excel y prescindiendo del uso de Power BI por el momento, el estudiante localiza e interpreta las variables utilizadas para calcular los indicadores presentes en las visualizaciones solicitadas, lo que contempla la segunda etapa de la metodología.

Paralelamente, utilizando el programa SQL Server Management Studio, se procede a buscar las variables presentes en el documento Excel en las tablas de Azure Synapse. La complejidad de esta búsqueda radica en que los nombres de las variables suelen diferir al de las existentes en el documento original y también en algunas ocasiones, los datos no están cargados de la misma manera que en el documento Excel. Esto último sucede cuando se requiere una metodología de cálculo para llegar a un dato comparable, o bien, porque las variables existentes una tabla de Synapse no son suficientes para llegar a los datos buscados.

Uno de los problemas más recurrentes en esta parte corresponde a que las tablas de Azure Synapse que deberían contener los datos para replicar las visualizaciones, no contienen la cantidad de información suficiente para establecer una comparación de los datos entre las distintas fuentes de origen. Esto generalmente se da en las etapas del ciclo productivo de pavos que, desde un punto de vista temporal son más recientes y, en consecuencia, llevan menos tiempo almacenando información.

Para solucionar estos problemas, existe la opción de obtener los datos faltantes desde otras tablas disponibles en Azure Synapse. Esto es posible, porque algunas tablas de hechos pueden compartir algunas variables, no obstante, corresponde a una solución transitoria para responder a las necesidades de monitoreo de indicadores prioritarios por los usuarios de Sopraval, pues no se está siguiendo el modelo lógico dimensional con el que fueron creadas las tablas, lo que dificulta el establecimiento de relaciones funcionales entre estas.

Adicionalmente, es necesario reunirse con el encargado de la carga de información en las tablas, quien tiene un entendimiento superior del ciclo productivo y puede dar alternativas en el caso de que no se pueda acceder a la variable de manera análoga al Excel. De esta forma, se levanta la alerta a los encargados de la recopilación, ingesta y almacenamiento de los datos para que se puedan encargar de estas falencias en el proceso de datos.

El tiempo que toma esta etapa depende de la prolijidad con que se hayan cargado los datos en las tablas y las dificultades que se tengan para encontrar

las variables. En este punto se encuentra una de las principales diferencias entre los dos tipos de tablas de hechos mencionados anteriormente. Para generar las visualizaciones en torno a la temática de "*Nacimiento*", correspondiente a una tabla de hechos del primer tipo, fue necesario buscar solo cinco variables. En cambio, en la temática de "*Cierres*", una tabla de hechos del segundo tipo, para generar las visualizaciones solicitadas fue necesario localizar alrededor de cuarenta variables, demorando más tiempo en esta parte y presentando una mayor cantidad de dificultades respecto a la información presente en las tablas de Azure Synapse.

Al ubicar estos datos, el estudiante agrupa las variables presentes en ambas fuentes de origen, en el caso de los datos presentes en Excel, estos son agrupados utilizando una tabla dinámica. Respecto a la información presente en las tablas de Azure Synapse, esta se agrupa mediante consultas en lenguaje SQL. El nivel de agrupación más común en esta labor corresponde al lote, ya que en todas las etapas del proceso de producción de pavos se identifica este nivel de agrupación en torno a fechas, lo que permite entender qué información fue cargada de manera más reciente e interpretar posibles diferencias que existan entre las diferentes fuentes de origen.

### 2.9.3 Comparación de datos agrupados y retroalimentación a personal de Sopraval

En este punto de desarrollo del proyecto se da lugar a la tercera etapa mencionada en la metodología, donde se incorporan los datos presentes en ambas fuentes de origen en una hoja de Excel. Así se crea una tabla comparativa, detallando desde donde se obtuvo la variable (en Excel, la hoja y en Synapse, la tabla) y bajo qué columna se accedió al dato buscado. Por cada grupo de datos, se verifica si los datos son idénticos entre ambas fuentes de origen. Esto se puede visualizar de la siguiente manera, aplicado a la temática "*Nacimiento*" mencionada anteriormente.

Hoja/Tabla	BD Nacimiento	NACIMIENTO	BD Nacimiento	NACIMIENTO
Nombre Variable	Total Nacidos	NACTOTAL	Huevos Cargados	CARGTRANSFERIDOS
Lote	Total Nacidos Excel	Total Nacidos Synapse	Huevos Cargados Excel	Huevos Cargados Synapse
21	51.937	51.937	61.651	61.651
20	51.976	51.976	61.651	61.651
19	50.815	50.815	61.641	61.641
18	52.767	52.767	61.651	61.651
17	50.850	50.850	61.567	61.567
16	52.922	52.922	61.646	61.646
15	51.260	51.260	61.641	61.641
14	52.747	223.750	61.520	260.583
13	53.678	593.519	61.649	682.021
12	235.910	1.031.390	287.827	1.209.853
11	520.748	1.186.781	609.401	1.406.332
10	848.020	1.507.303	974.091	1.782.270
9	806.700	1.219.866	956.498	1.501.454
8	1.460.279	1.925.708	1.768.105	2.423.179
7	1.466.086	1.604.675	1.886.660	2.124.240
6	1.138.549	1.148.590	1.525.978	1.555.974
5	1.515.844	1.499.174	2.026.882	2.006.430
4	1.713.115	1.687.744	2.109.326	2.078.648
3	1.727.246	1.606.710	2.085.757	1.932.370
2	1.658.484	1.318.113	2.080.903	1.676.984
1	1.907.025	1.641.863	2.321.019	2.001.293

Ilustración 8: Tabla comparativa nacimiento (Fuente: Elaboración propia).

En la tabla comparativa, se muestra que, para los datos de los lotes más recientes, correspondientes a los lotes 15 hasta el 21 los valores son idénticos, por ende, la información fue cargada correctamente en la nueva fuente de origen y se marcan en verde. Para los lotes 6 hasta el 14, los valores de los datos en Azure Synapse son mayores a los del Excel, por lo que son marcados en amarillo. Por último, para los datos entre el lote 1 hasta el 5, los datos de la nueva fuente de origen son menores a los del documento Excel, por lo que son marcados en rojo. Tanto para los valores marcados en amarillo y en rojo, se tienen diferencias que no deberían existir y que hay que corregir, dando lugar a la componente de validación de los datos con personal de Sopraval que se explicará enseguida. Es importante destacar que los valores mostrados en esta tabla no corresponden a los datos reales de la empresa, pues su único fin es explicitar la metodología empleada para la comparación.

Posterior a la comparación entre las variables existentes en ambas fuentes de origen, comienza la retroalimentación de los datos presentes en las tablas de Azure Synapse. Para esto, es necesario contactar al responsable de la carga de los datos en estas tablas, que corresponde a un empleado de Sopraval, y explicitar qué grupos de datos no están coincidiendo en la comparación, otorgándole la tabla comparativa previamente generada.

Sumado a lo anterior, es fundamental exponer a la contraparte el procedimiento mediante el que se accedió a las variables en las distintas fuentes de origen y que aprueben las metodologías mediante las que se hayan calculado las variables en caso de que sean diferentes a las del Excel original. Esto sirve para asegurar que la lógica utilizada por el estudiante a lo largo del proceso de búsqueda sea la correcta, además de reportar a los responsables de la recopilación, ingesta y almacenamiento de los datos, que deben corregir las diferencias existentes en la comparación, entendiendo los motivos por los que se cargaron mal los datos en la nueva fuente de origen. Estas deficiencias en la información vienen desde niveles anteriores del proceso de datos y no son responsabilidad del equipo de Analítica y BI del área de Producción Animal.

## 2.9.4 Carga de información en Power BI y establecimiento del modelo dimensional

Una vez el equipo reporta la existencia de errores en la carga de información con el personal de Sopraval, se comienzan a cargar las tablas de Azure Synapse a un nuevo documento Power BI, considerando que los responsables de los niveles anteriores del proceso de datos ya se están encargando de corregir las diferencias explicitadas en la comparación. En esta parte es necesario comprender el modelo lógico mediante el que se relacionan las tablas, entendiendo cuáles de ellas son tablas de dimensión y cuáles son tablas de hechos.

Como se mencionó anteriormente, las tablas de hechos están ligadas a una temática en particular y contienen las variables necesarias para generar las visualizaciones de dicha temática. Estas tablas poseen los datos que se quieren medir y analizar en una fase posterior, por lo que en general tienen muchos registros. Las tablas de dimensión, por su parte, tienen una menor cantidad de registros y se conectan con las tablas de hechos para aportar con las perspectivas desde la que se quieren medir y analizar los datos.

En general, las tablas de dimensión utilizadas suelen ser transversales para las distintas temáticas, salvo unos casos más específicos en el que se busca analizar la información desde otras perspectivas puntuales. Las tablas de dimensión más utilizadas a lo largo de los modelamientos corresponden a "*Proceso*", "*Etapa*", "*Granja*" y "*Pabellón*". Respectivamente, estas tablas indican las dimensiones de proceso, etapa, granja y pabellón desde el que el dato fue extraído y aportan en su conjunto a la temporalidad y ubicación física de origen de los datos. Para que las tablas dimensionales aporten a la segmentación de la información, es necesario establecer correctamente las relaciones entre las tablas de Azure Synapse. En caso de no seguir la estructura de su modelo lógico, los objetos visuales a replicar presentarán diferencias respecto a los

objetos presentes en el Power BI original al disponer las visualizaciones desde alguna de las perspectivas que aportan tablas dimensionales.

Entendiendo esta lógica, se cargan todas las tablas que, en su conjunto, contienen la información necesaria para replicar una visualización y se relacionan de acuerdo con la estructura del modelo dimensional.

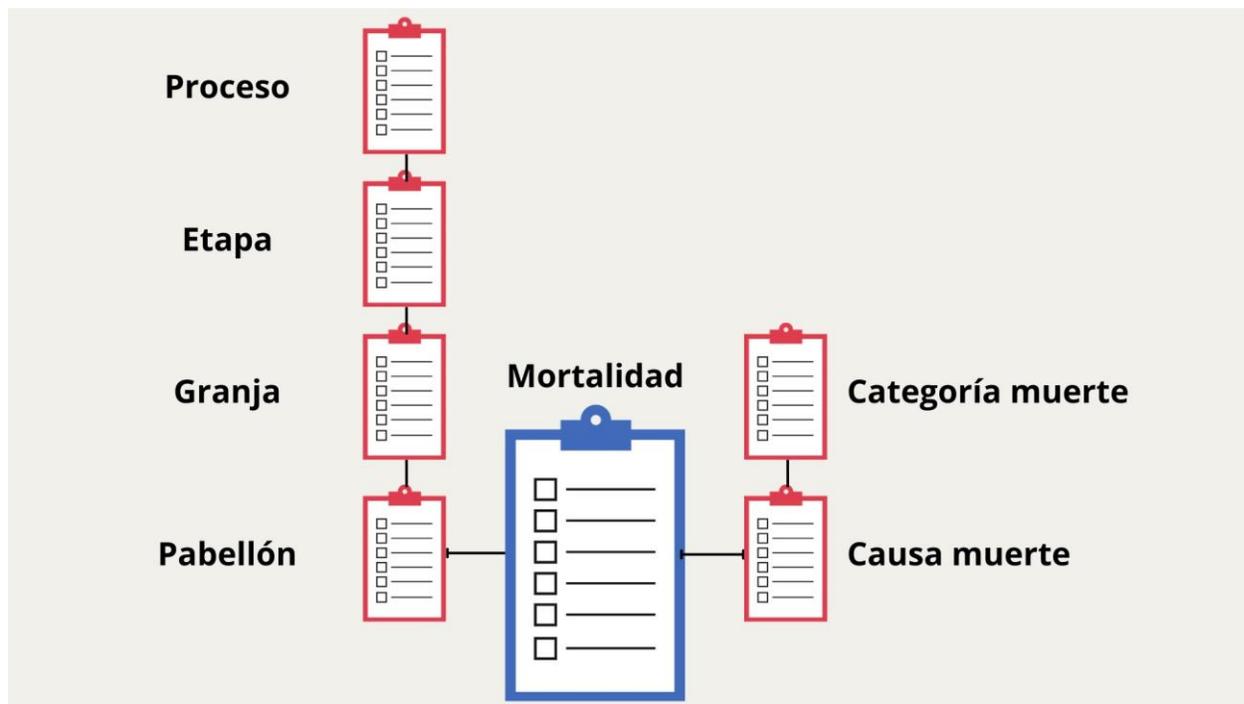


Ilustración 9: Ejemplo estructura dimensional de la tabla Mortalidad (Fuente: Elaboración propia).

En este ejemplo, se muestran algunas de las tablas dimensionales relacionadas a la tabla de hechos "Mortalidad" y cómo se conectan entre sí. Esta estructura permite segmentar los datos existentes en la tabla de hechos mediante las dimensiones existentes en cada tabla dimensional. Por ejemplo, la tabla dimensional "Causa muerte" aporta con la perspectiva de causa por la que muere un animal. Considerando que existe una causa llamada "Infarto" en esa tabla, entonces en una visualización se podría segmentar la información de la tabla "Mortalidad", de modo que solo se muestren datos cuya causa de muerte de los animales fue provocada por un infarto.

Desde que las tablas pertenecientes a la nueva fuente de origen son cargadas a Power BI, las correcciones de las diferencias en los datos en las etapas de recopilación, ingesta y almacenamiento del proceso de datos se verán reflejadas también en las visualizaciones, cumpliendo con la primera parte del objetivo general, pues el proceso de datos se encuentra automatizado. Hasta este

punto del desarrollo del proyecto, el estudiante es el principal responsable de todas las tareas recientemente expuestas. El aporte que han realizado sus compañeros del equipo de Analítica y BI se concentra en el desarrollo de dashboards, que será expuesto a continuación.

Una vez cargadas las tablas de Azure Synapse en Power BI, se conectan siguiendo la estructura del modelo Snowflake, referido en el marco conceptual y la metodología. Para realizar las conexiones, en algunas ocasiones es necesario utilizar Power Query, que corresponde a una herramienta dentro de la plataforma Power BI que permite realizar transformaciones a los datos (en este caso a las tablas de Azure Synapse) para poder trabajar con ellos, sin embargo, el modelamiento de los datos está pensado para que las tablas sean conectadas siguiendo la estructura del modelo dimensional, por lo que corresponden a situaciones puntuales y no se profundizará al respecto.

Las relaciones entre las tablas son establecidas manualmente, utilizando un tipo de relación de "uno a varios". Esto significa que un dato, que está registrado en la tabla dimensional una sola vez, se conecta a la tabla de hechos mediante la variable que contiene el mismo dato, donde los datos se encuentran registrados en muchas instancias. Entre tablas de dimensiones ocurre análogamente, pues el tipo de relación es el mismo, aplicado para dos tablas dimensionales.

Siguiendo el ejemplo anterior, la causa "Infarto" está registrada en la tabla "Causa Muerte" en una sola ocasión. En la tabla "Mortalidad", también existe la causa "Infarto", pero registrado en múltiples ocasiones, pues muchas de las muertes registradas se dieron por tal causa. Entonces la relación entre estas dos tablas se realiza desde la tabla "Causa muerte" hacia "Mortalidad", conectando las tablas mediante una relación de "uno a varios".

## 2.9.5 Desarrollo de las nuevas visualizaciones en Power BI

Una vez efectuadas las relaciones entre todas las tablas necesarias para generar las visualizaciones, los integrantes del equipo de Analítica y BI se dedican a replicar los objetos visuales en Power BI, dando espacio a la quinta etapa de la metodología. En esta parte del desarrollo, es relevante tener conocimientos en lenguaje DAX (el lenguaje utilizado para realizar los cálculos en Power BI), porque para reproducir algunos objetos visuales es necesario aplicar cálculos inexistentes en el documento Power BI original, por lo que la metodología de replicar las visualizaciones no es directa. Esto último se debe a que, a pesar de que las distintas fuentes de origen tienen el potencial de aportar con la misma información necesaria para generar la visualización, a nivel estructural

la tabla puede diferir en la manera que se disponen los datos, o bien, los datos no existen en las tablas de Azure Synapse de la misma manera que estaban en la fuente de origen anterior, lo que se traduce en la aplicación de cálculos adicionales para llegar a obtener un mismo indicador.

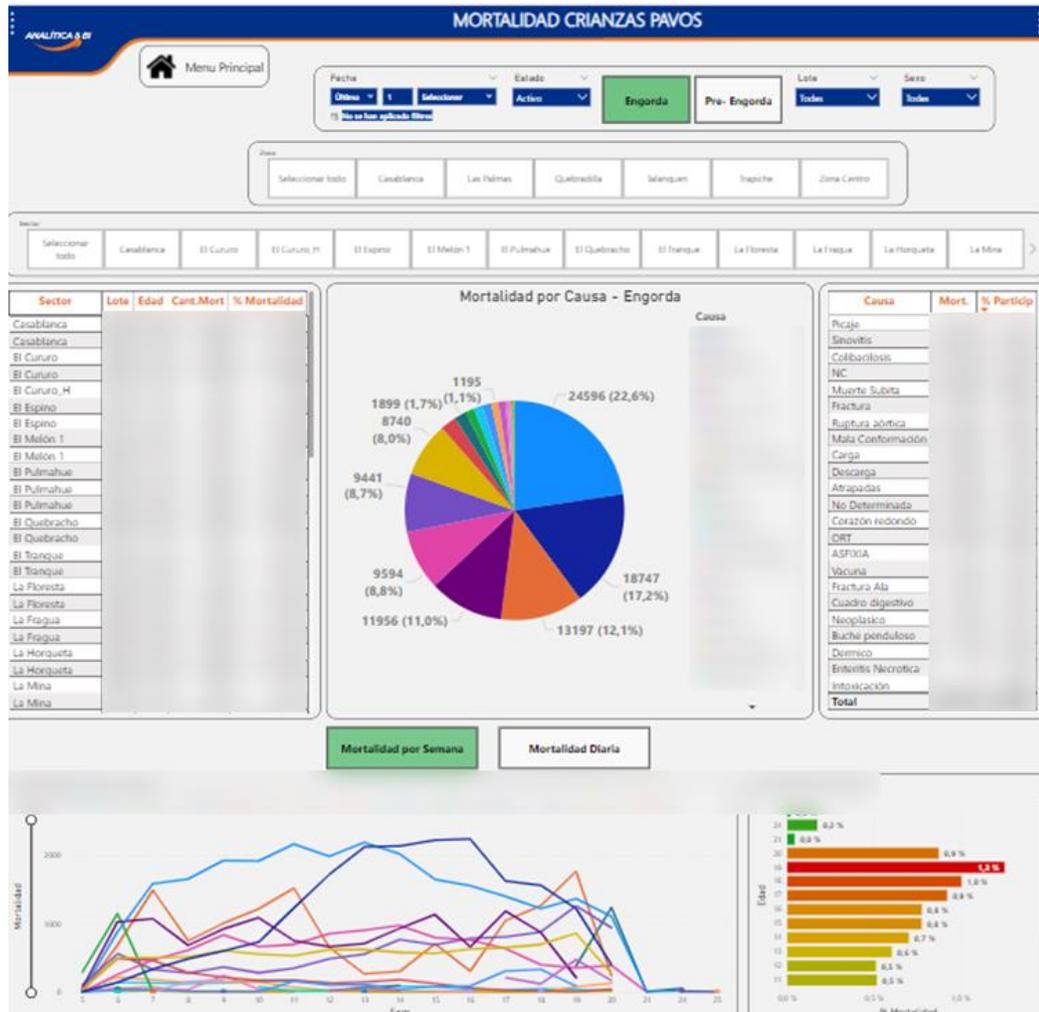


Ilustración 10: Panel de temática Mortalidad (Fuente: Elaboración Analítica y BI de Producción animal).

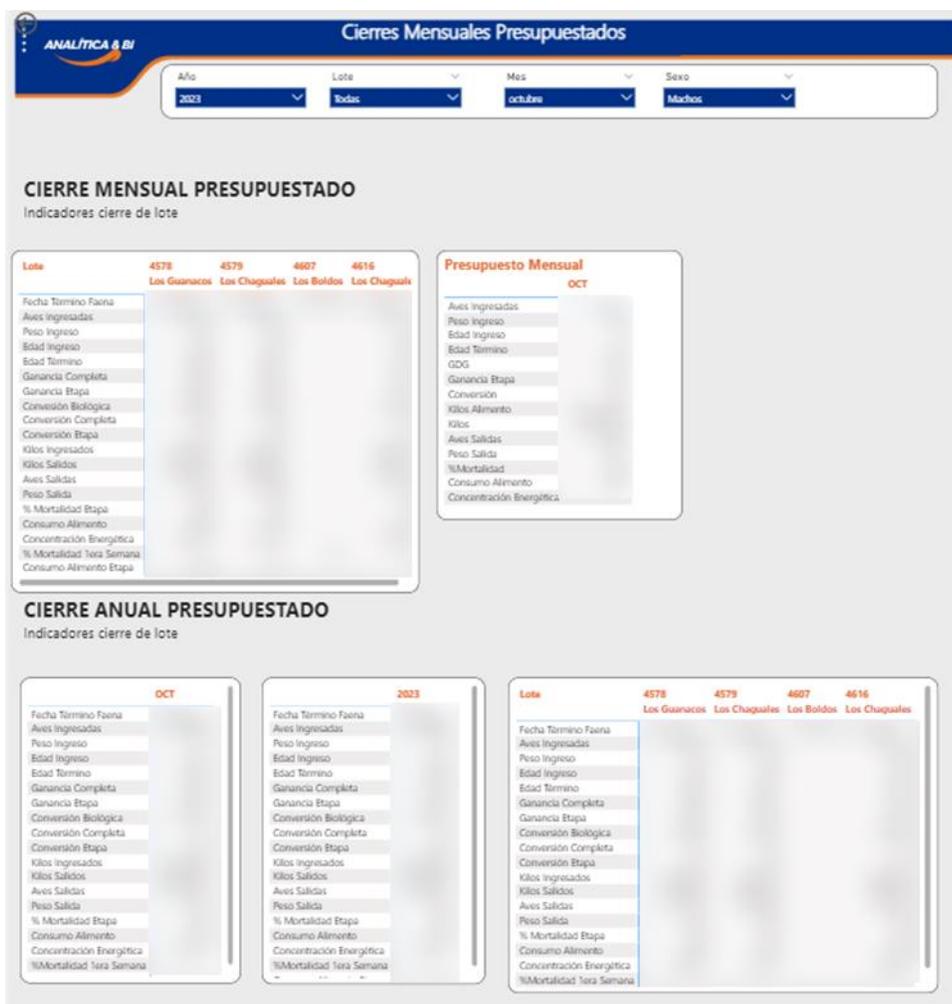


Ilustración 11: Panel de temática Cierres (Fuente: Elaboración Analítica y BI de Producción animal).

Abordando los dos tipos de tablas de hechos mencionados al comienzo de esta sección, la complejidad difiere entre ambos casos en el desarrollo de los paneles. De manera inversa a lo que ocurría localizando variables entre fuentes de origen, en el caso de generar visualizaciones, la complejidad ha aumentado en las tablas que no poseen los indicadores calculados. Esto se debe principalmente a que estas tablas abarcan una menor cantidad de indicadores, por lo que las dimensiones desde las que se aborda un indicador para generar las visualizaciones pueden ser más específicas que al abordar muchos indicadores en conjunto. En estos casos se genera una mayor cantidad de objetos visuales, pues se abordan los indicadores desde más perspectivas, y algunos de ellos poseen una complejidad superior durante el proceso de replicación. Este corresponde al caso del panel de "Mortalidad", en donde se aprecia los indicadores de "Mortalidad total" y "% Mortalidad", con varios objetos visuales distintos en el mismo panel.

En el caso de las tablas de hechos que contienen los indicadores previamente calculados, en la mayoría de las visualizaciones solicitadas no es necesario aplicar cálculos a los datos cargados desde Azure Synapse, pues estos cálculos fueron efectuados y cargados previamente a dicha tabla de hechos. También al abarcar muchos indicadores de temáticas diferentes, estos suelen mostrarse en objetos visuales más simples y no es posible mostrar los indicadores desde muchas perspectivas diferentes. Esto último se aprecia en la visualización de la temática de "Cierres", donde se aprecian diecinueve indicadores diferentes, pero limitados a una perspectiva mensual y anual, mostrados en objetos visuales más simples.

Un aspecto importante durante el desarrollo de las vistas es que, al momento de generar los objetos visuales, se debe verificar que las visualizaciones del Power BI original y las replicadas sean idénticas para los grupos de datos que cuadraban al efectuar la comparación entre ambas fuentes de origen, por lo que las tablas comparativas generadas para la validación también son fundamentales en esta instancia del proyecto. Para los datos que no cuadraban cuando se hizo la tabla comparativa, se tiene que a medida que se vayan actualizando y arreglando las diferencias en el proceso de datos, los cambios también se verán reflejados en las visualizaciones, por lo que, si las visualizaciones son idénticas para los lotes que cuadraban, también lo serán para los grupos de datos que aún no se han corregido, dando por finalizada la quinta etapa de la metodología.

## 2.9.6 Validación y publicación de las nuevas visualizaciones

Una vez las visualizaciones son replicadas, estas son revisadas y validadas por el jefe de Analítica y BI, el que luego de solicitar modificaciones en las nuevas vistas generadas y revisar detalles estéticos, se encarga de publicar las nuevas visualizaciones, lo que permite al personal de Sopraval acceder al nuevo documento Power BI. La frecuencia con la que debe ser actualizada el panel depende de los indicadores que se busque monitorear y de los usuarios finales de los paneles obtenidos, por lo que es el jefe del equipo quien se encarga de contactar a estos usuarios para acordar la tasa de actualización del panel.

Al estar publicado el nuevo documento Power BI, posibilitando el acceso al personal de Sopraval y finalizando la última etapa de la metodología, se termina un ciclo del proyecto ligado a una de las temáticas abordadas. Por lo que se agenda una nueva reunión con el personal de Sopraval, quienes solicitan nuevas visualizaciones a replicar, repitiendo la metodología expuesta.

## 2.9.7 Aspectos transversales

En las diferentes instancias del proyecto, el estudiante documenta el desarrollo de este aplicado a cada una de las temáticas para procurar la mejora continua de las nuevas visualizaciones y poder modificarlas a medida que se generen nuevas necesidades por parte de los usuarios. La documentación sirve tanto para el alumno como para quienes continúen el desarrollo del proyecto en un futuro, pues el manejo de la totalidad de las tablas y la información que contienen requiere un proceso de adaptación y conocer los indicadores del proceso en profundidad. Asimismo, esto proporciona una herramienta que apoye especialmente en la localización y comparación entre las diferentes fuentes de origen.

Además, a lo largo de las distintas etapas el estudiante se encarga de proponer alternativas que puedan aportar a la eficiencia y efectividad del proceso de datos y del desarrollo de paneles al resto del equipo de Analítica y BI. Para encargarse de este punto, el estudiante completó el curso avanzado de Power BI y ha investigado sobre el proceso de datos, lo que le ha permitido generar algunas de las nuevas visualizaciones por su cuenta (aunque la mayoría han sido realizadas por el resto del equipo) y entender cómo utilizar adecuadamente los datos de la nueva fuente de origen, siguiendo buenas prácticas a lo largo del proyecto y contrastando con algunas de las herramientas empleadas a lo largo del desarrollo.

## 2.10. Resultados

En las cinco temáticas mencionadas al comienzo de la sección del anterior, se han completado los requerimientos e indicadores prioritarios asociados a cada una de ellas. A pesar de esto último, se debe destacar que cada temática será abordada nuevamente a medida que vaya avanzando el proyecto, en base a las nuevas solicitudes que surjan por parte de los usuarios de los nuevos paneles e incorporando nuevos indicadores de menor relevancia que los ya abordados.



*Ilustración 12: Paneles abordados por temática (Fuente: Elaboración propia).*

Como se explicó en el desarrollo del proyecto, dependiendo de las temáticas y sus tablas de hechos correspondientes, se generaron dificultades en distintas etapas del proceso. No obstante, todas las visualizaciones han podido ser generadas a cabalidad desde los datos existentes en las tablas de la nueva fuente de origen. En consecuencia, se han replicado un total de veintiún paneles, teniendo la capacidad de responder a todas las solicitudes y cumpliendo cada uno de los objetivos específicos, lo que implica a su vez el cumplimiento del objetivo general en todas las iteraciones del desarrollo del proyecto.

El cumplimiento de los primeros dos objetivos específicos permite prescindir de la labor de carga manual de los datos en los paneles abordados, en donde solo considerando las temáticas respectivas al "Nacimiento" y "Postura y carga de huevos", el encargado de realizar dicha labor estima que antes del proceso

de conexión a la nueva fuente de origen demoraba alrededor de dos horas semanales para mantener dichas visualizaciones actualizadas. Considerando la situación actual en que solo se está utilizando la nueva fuente de origen en esos paneles dicha tarea ya no es necesaria, ya que el proceso de datos se encuentra automatizado.

En relación con el tercer objetivo específico, como resultado del desarrollo de los paneles de desempeño, se tiene que las visualizaciones para el monitoreo de indicadores se actualizan, en general, veinte veces más cada semana respecto a los paneles anteriores. Esto se debe a que un panel que se actualizaba una vez a la semana, ahora se está actualizando tres veces por día. Lo anterior, permite reaccionar adecuadamente a indicadores críticos para el desarrollo del negocio, aportando a la detección de incidentes de manera oportuna y otorgando información actualizada para tomar mejores decisiones. Un ejemplo reciente al respecto corresponde a un incidente vinculado a la influenza aviar y al virus ISA, que resultó en pérdidas de 85 millones de dólares para Agrosuper (Baeza M. 2023), donde mejoras en el proceso de datos y paneles de mortalidad actualizados con mayor frecuencia hubieran aportado a la prevención de una situación tan masiva como la expuesta.

Para ejemplificar el resultado final de los paneles desarrollados, se mostrará un antes y después para una visualización de la temática "Nacimiento", generadas desde las fuentes de origen respectivas.

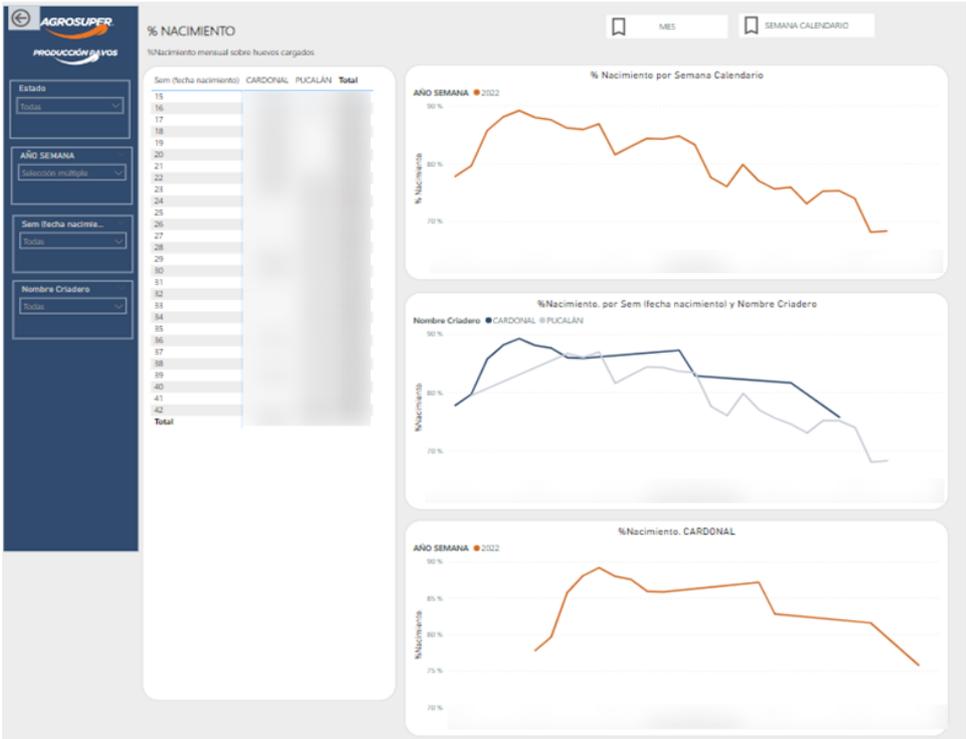


Ilustración 13: Panel nacimiento a replicar (Fuente: Elaboración Analítica y BI de Producción animal).

En la presente imagen, se muestran los objetos visuales originales, es decir los que tienen como fuente de origen un Excel cargado de forma manual una vez a la semana. En la siguiente imagen, se muestran los objetos visuales replicados, cuya fuente de origen viene de un proceso de datos automatizado y orquestado según el concepto de data pipeline, con mayor frecuencia de actualización.



Ilustración 14: Panel nacimiento replicado (Fuente: Elaboración Analítica y BI de Producción animal).

Es importante notar que las curvas en los gráficos son idénticas para ambas visualizaciones, lo que se debe a que se utilizó un ejemplo de agrupación en el que los datos son los mismos, a pesar de venir desde fuentes de origen distintas. Esto quiere decir, que en el proceso de comparación los datos utilizados para generar estos objetos visuales fueron marcados en verde, pues cuadraban entre ambas fuentes de origen. Las principales diferencias en ambos paneles son estéticas, pero las funcionalidades que se tienen y las perspectivas desde las que se busca medir y monitorear el indicador son las mismas. Por lo que el valor no está en las modificaciones que se le agreguen a las visualizaciones, sino que está en la manera que estos paneles acceden a los datos mediante la carga de información y la frecuencia con que estos paneles se actualizan.

En relación con el cuarto objetivo específico, el estudiante ha aportado al equipo proponiendo otras metodologías para realizar una misma tarea, esto principalmente en la etapa de carga de datos de la nueva fuente de origen en Power BI. Donde existen oportunidades de mejora en la cantidad de datos que se cargan para replicar los paneles, desde la perspectiva de que, en lugar de cargar todas las tablas y establecer las relaciones entre estas dentro de la plataforma BI, es posible generar transformaciones en los datos que permitan cargar una menor cantidad de información en Power BI y disminuir la cantidad de cálculos efectuados en la plataforma para replicar un panel, mejorando el rendimiento de esta y eludiendo algunas de las dificultades que se presentan durante el desarrollo de los paneles.

Finalmente, el estudiante ha logrado generar una documentación de todo el proceso de desarrollo del proyecto, la que puede ser utilizada por quién se encargue de continuar el desarrollo de este a futuro, mostrando las metodologías utilizadas en conjunto con los archivos generados en las etapas anteriores al desarrollo de los paneles. Lo anterior, facilita principalmente el entendimiento del modelo lógico de las tablas en conjunto con la búsqueda de la información que aporta cada tabla de Azure Synapse.

## 3. Conclusiones

En base a la ejecución del proyecto mediante la metodología previamente expuesta, se ha replicado la totalidad de los paneles solicitados por Sopraval en las cinco temáticas abordadas, cumpliendo con cada uno de los objetivos específicos planteados en todas las iteraciones del proyecto. Los resultados obtenidos, permiten a los encargados de monitorear los indicadores utilizar paneles que se actualizan constantemente mediante información que proviene un proceso de datos automatizado.

La automatización de este proceso de actualización de paneles conlleva diversas ventajas. En primer lugar, favorece la toma de decisiones y mejora la capacidad de respuesta a incidentes al proporcionar información actualizada en tiempo real. Lo anterior, se debe a que los indicadores presentes en los paneles se generan a partir de los datos más recientes almacenados en el momento de la actualización, lo que ofrece a los responsables una visión más completa y precisa de la situación en comparación con los paneles originales. En segundo lugar, automatizar elimina la necesidad de carga manual de datos en los paneles por parte de un Ingeniero del equipo de Analítica y BI. Esta mejora no solo ahorra tiempo, sino que también previene la existencia de errores durante el proceso de carga de información en los paneles.

Los desafíos inherentes al proceso de desarrollo del proyecto se centran en dos temas principales: la validación de la cuadratura de los datos entre ambas fuentes de origen, además de la eficiencia y efectividad de las metodologías empleadas a lo largo del proceso. Respecto a estos temas, se ha dado lugar a problemáticas recurrentes en las iteraciones del proyecto que se explicarán respectivamente a continuación.

Una de las problemáticas corresponde a la carga errática de datos en las tablas de Azure Synapse, donde se presentan discrepancias entre la nueva fuente de origen de datos y la fuente anterior, obstaculizando el progreso del proyecto. Estas diferencias son consecuencia de etapas anteriores en el proceso de datos, las que deben ser esporádicamente reportadas a sus encargados para ser corregidas y escapan de las responsabilidades del equipo al que pertenece el estudiante. Si se mejorara la prolijidad con la que se cargan los datos, entonces el enfoque de los Ingenieros del equipo respecto al proyecto estaría principalmente en desarrollar nuevos paneles más que en validar que los datos cuadren, lo que aumentaría su capacidad de responder a nuevas solicitudes de la contraparte y mejoraría la fluidez en el avance del proyecto.

Otra problemática existente radica en la resistencia a la posibilidad de generar consultas SQL en la etapa de modelamiento de los datos por parte del equipo de Analítica y BI, alternativa que permitiría crear una tabla única que contenga solo la información necesaria para replicar un panel y prescindir de la carga de varias tablas para generar el mismo. El motivo principal de esta reticencia está ligado a la falta de personal en el equipo de Arquitectura de Datos para desarrollar nuevas consultas. No obstante, si se delegara la generación de consultas al equipo del estudiante, con validación por parte del equipo de Arquitectura de Datos, podría superarse esta limitante, mejorando la eficiencia del documento BI generado a partir de la nueva fuente de origen. La metodología actual, que implica cargar muchas tablas con datos innecesarios, conduce a ralentizaciones y requiere cálculos adicionales para adaptarse a la disposición de los datos en nuevas tablas para replicar los objetos visuales.

Sumado a lo anterior, durante el desarrollo del proyecto se ha incurrido a prácticas ineficientes para cumplir con la urgencia de los requerimientos, como por ejemplo cargar tablas adicionales en visualizaciones que no deberían requerirlas para replicar un panel específico o realizar transformaciones a los datos dentro de la plataforma Power BI. Estos son ejemplos de prácticas que pueden incidir negativamente en el rendimiento de los nuevos paneles, aún más considerando que se seguirán incorporando nuevos datos en la nueva fuente de origen. Estas prácticas podrían ser abordadas integralmente abriéndose a la metodología del párrafo anterior.

Un último aspecto que dificultó el desarrollo del proyecto corresponde a la semántica de los datos, en donde los nombres de las variables pertenecientes a las tablas de Azure Synapse suelen ser difíciles de interpretar, incluso para el personal de Sopraval. La inexistencia de un diccionario que describa en detalle el significado de las tablas y sus variables dificulta el proceso de búsqueda, sobre todo pensando que quienes realizan las visualizaciones corresponden a un equipo externo a Sopraval, llevando a equivocaciones difíciles de detectar.

Más allá de los problemas recientemente expuestos, Agrosuper es una empresa con buena cultura de datos, ya que al tener una cadena de valor verticalmente integrada controla y mantiene la calidad en sus operaciones en las diferentes componentes del negocio a partir del monitoreo de indicadores. Particularmente, en el área de Producción Animal, se presentan desafíos aún mayores para mantenerse a la vanguardia en tecnologías de la Transformación Digital. Por ejemplo, en los indicadores pertenecientes a la temática de "Mortalidad", la relevancia del proyecto abordado es aún mayor, ya que el actualizar los paneles veinte veces más cada semana respecto a la situación anterior permite amortiguar las repercusiones producidas por una situación como la

acontecida a causa de la influenza aviar, lo que a un nivel masivo puede ser de carácter catastrófico para la empresa.

La automatización completa del ciclo productivo de pavos, desde la recopilación hasta la carga de datos en Power BI, emerge como un hito significativo. Este proceso, único entre todas las áreas de la empresa, establece un precedente para la transformación digital en otros departamentos. Hasta ahora, el proyecto ha validado la viabilidad de esta perspectiva, revelando una herramienta con el potencial de generar valor en todas las áreas del negocio.

En este contexto, la documentación generada por el estudiante adquiere una relevancia que se extiende más allá del ámbito específico del ciclo de pavos. Los conocimientos adquiridos durante el desarrollo de este proyecto no solo beneficiarán su desempeño como profesional en este campo, sino que también aportarán valiosas lecciones aplicables a otras áreas dentro de la empresa. Este enfoque integral subraya la importancia estratégica de la automatización y cómo su implementación puede redefinir la eficiencia operativa en diversos aspectos de la organización.

# Bibliografía

- (1) Agrosuper. (2023). Reporte Integrado Agrosuper 2022. <https://www.agrosuper.cl/sustentabilidad/reportes/>
- (2) Raman, K., Swaminathan, A., Gehrke, J., & Joachims, T. (2013, August). Beyond myopic inference in big data pipelines. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 86-94).
- (3) Miloslavskaya, N., & Tolstoy, A. (2016). Big data, fast data and data lake concepts. *Procedia Computer Science*, 88, 300-305.
- (4) Ballard, C., Farrell, D. M., Gupta, A., Mazuela, C., & Vohnik, S. (2012). *Dimensional Modeling: In a Business Intelligence Environment*. IBM Redbooks.
- (5) Elena, C. (2011). Business intelligence. *Journal of Knowledge Management, Economics and Information Technology*, 1(2), 1-12.
- (6) Eckerson, W. W. (2010). *Performance dashboards: measuring, monitoring, and managing your business*. John Wiley & Sons.
- (7) Sadiku, M., Shadare, A. E., Musa, S. M., Akujuobi, C. M., & Perry, R. (2016). Data visualization. *International Journal of Engineering Research And Advanced Technology (IJERAT)*, 2(12), 11-16.
- (8) Domínguez, E., Pérez, B., Rubio, Á. L., & Zapata, M. A. (2019). A taxonomy for key performance indicators management. *Computer Standards & Interfaces*, 64, 24-40.
- (9) Baeza, M. (4 de agosto de 2023). Agrosuper registra pérdidas por US\$ 85 millones con peores desempeños en carnes y salmones. *Diario Financiero*. <https://www.df.cl/empresas/industria/agrosuper-registra-perdidas-por-us-85-millones-con-peores-desempenos-en>

# Anexo A

## A.1. Certificaciones segmento Carnes

**PRINCIPALES CERTIFICACIONES**

 **SIGAS**  
Sistema Integrado de Gestión de Agrosuper (SIGAS), el cual se integra y se potencia con otras metodologías implementadas como es OPI (Optimización de Procesos Industriales, basado en TPM) que nos permiten trabajar de manera transversal con las distintas unidades de negocio, apuntando a una mirada global de la calidad, donde cada uno conoce sus responsabilidades en el proceso y resultado.

 **HACCP**  
Sistema que aborda la seguridad alimentaria a través de la identificación, análisis y control de los peligros físicos, químicos, biológicos y radiológicos. Abarca desde las materias primas y las etapas del proceso de elaboración, hasta la distribución y consumo del producto terminado.

 **BRGS**  
Esquema diseñado para consolidar estándares de seguridad alimentaria en toda la cadena de suministro, reconocido mundialmente en las industrias alimentarias y no alimentarias como uno de los esquemas de certificación de terceros más rigurosos.

 **ISO** International Organization for Standardization  
**ISO50001**  
La certificación ISO50001 demuestra el compromiso de una organización con la mejora continua de la gestión de la energía, garantizando el cumplimiento de requisitos legislativos y reglamentarios correspondientes.

Anexo A.1.: Certificaciones segmento Carnes (Fuente: Reporte Integrado Agrosuper 2022).