

**DASHBOARD PARA LA VISUALIZACIÓN DE INDICADORES DE RATING Y
RANKING DE LOS CANALES DE TELEVISIÓN COMERCIALIZADOS EN
VILLAVICENCIO.**

Autores

Manuel Alejandro Murcia Sánchez
Geovani Leandro Martínez Barón

Director

Elio H. Cables Pérez, Ph.D

Universidad Antonio Nariño
Facultad de Ingeniería de Sistemas
Especialización en Gobierno de Datos
Bogotá D.C
2023

Tabla de contenido

1	Introducción	1
2	Descripción y formulación del problema	2
2.1	Formulación del problema.....	3
2.1.1	Objetivo General.....	3
2.1.2	Objetivos Específicos	3
3	Marco referencial	4
3.1	Marco teórico.....	4
3.1.1	Canales de televisión:	5
3.1.2	Decodificador:	5
3.1.3	Rating:	5
3.1.4	Ranking:	6
3.1.5	Addons:	6
3.1.6	Recopilación de orígenes de datos:	6
3.1.7	Franjas Horarias:	7
3.1.8	Bases de datos:.....	7
3.1.9	Sistemas de administración de bases de datos (SGBD):	7
3.1.10	Modelo de datos:	8
3.1.11	Proceso de Extract, Transform, Load (ETL):	9
3.1.12	Herramientas de ETL:	10
3.1.13	Calidad de datos:	10
3.1.14	Limpieza de datos:.....	12
3.1.15	Base de datos Staging:	12
3.1.16	Herramientas de visualización:	12
3.2	Estado del arte	13
3.2.1	Kantar IBOPE Media:	13
3.2.2	Nielsen:.....	14
3.2.2	ComScore:.....	14
3.2.3	Tabla de resumen comparativo propuesta para Mayha vs empresas en el mercado:	14

3.3	Impacto	15
3.4	Componente de Innovación.....	16
4	Metodología.....	17
4.1	Identificación de datos.....	17
4.2	Consulta de los datos.....	17
4.3	Alistamiento de los datos.....	17
4.4	Almacenamiento	17
4.5	Procesamiento.....	17
4.6	Resultado.....	17
5	Desarrollo de la propuesta	19
5.1	Identificación de Datos.....	19
5.2	EPG.....	20
5.3	Event.....	21
5.4	Devices	21
5.5	Consulta de los datos:.....	22
5.6	Alistamiento de los datos:.....	33
5.7	Almacenamiento:	37
5.8	Procesamiento:.....	42
6	Resultado	53
7	Conclusiones.....	62
8	Referencias	63

Índice de Figuras

Ilustración 1. Ejemplo de arquitectura de datos corporativa tradicional.	10
Ilustración 2. ETL reglas de limpieza hacia Staging.....	12
Ilustración 3. Fases ejecutadas para el proyecto.	18
Ilustración 4. Entrega de Datos realizada por Mayha	19
Ilustración 5. Resultset de la fuente entregada por Mayha de EPG.	20
Ilustración 6. Resultset TOP 10 de la fuente entregada por Mayha tabla events. 21	
Ilustración 7. Resultset fuente entregada por Mayha denominada InventoryDevice.	22
Ilustración 8. Creación del grupo de recursos en Azure.	23
Ilustración 9. Creación del servidor SQL.....	23
Ilustración 10. Creación de base de datos Rating_EPG_Base_01 en nube con marcado indicado en la política Nro. 2.	24
Ilustración 11. Creación de base de datos Rating_Devices_Base_02 en nube con marcado indicado en la política Nro. 2.	25
Ilustración 12. Creación de base de datos Rating_Events_Base_03 en nube con marcado indicado en la política Nro. 2.	26
Ilustración 13. Muestra del Ambiente con las BD creados.	27
Ilustración 14. Importe del Archivo Device.bacpac a Rating_Device_Base_02....	28
Ilustración 15. Importe del Archivo EPG.bacpac a Rating_EPG_Base_01.	28
Ilustración 16. Importe del Archivo Events.bacpac a Rating_Evens_Base_03.	28
Ilustración 17. Resultado Exitoso de la importación.	29
Ilustración 18. Modelo lógico de base de datos EPG	30
Ilustración 19. Modelo Físico.	31
Ilustración 20. Tabla Events.....	32
Ilustración 21. Tabla InventoryDevice.	32
Ilustración 22. Tabla EPGHistorico sin relación.	33
Ilustración 23. Creación de restricciones.....	34
Ilustración 24. Creación de restricciones programas.	35
Ilustración 25. Consulta de verificación de existencia de programas de infomerciales.	36
Ilustración 26. Verificación de la existencia de programas emitidos con duración inferior a 5 minutos.	37
Ilustración 27. Desfragmentación de la fecha hora inicial y final de los programa y actualización en la base extraída.	38
Ilustración 28. Consulta que relaciona los eventos vs los programas.	39
Ilustración 29. Creación de Etl y actividad Copiar Datos.	39
Ilustración 30. Creación de objeto destino.	40
Ilustración 31. Configuración de Tabla destino EventTuning.....	40
Ilustración 32. Evidencia de procesamiento de ETL.	41
Ilustración 33. Resultado de la ETL CP_Up_EventTuning.	42
Ilustración 34. Creación de bd Rating_Staging_Base_04.	43

Ilustración 35. Creación de ETL PL_Etl_Info_01.	44
Ilustración 36. Configuración de ETL PL_Etl_Info_01.	44
Ilustración 37. Creación de linked server a bd Rating_EPG_Base_01.	45
Ilustración 38. Configuración de la actividad CP_Etl_Sin_Infomerciales_01.	46
Ilustración 39. Creación de linked server a bd Rating_Staging_Base_04.	47
Ilustración 40. Configuración destino tabla SabanaReporte.	47
Ilustración 41. Monitoreo sobre el ETL PL_Etl_Info_01.	48
Ilustración 42. Finalización exitosa del ETL PL_Etl_Info_01.	49
Ilustración 43. Evidenciade la inserción sobre la tabla SabanaReporte.....	49
Ilustración 44. Modelo estrella para dashboard de rating.	50
Ilustración 45. Formula en DAX cálculo del indicador	51
Ilustración 46. Formula en DAX calculo indicador share	51
Ilustración 47. Formula rating para franja.	51
Ilustración 48. Formula share para franja.....	51
Ilustración 49. Formula DAX cálculo de rating para programas	52
Ilustración 50. Formula DAX para cálculo de Share para programas	52
Ilustración 51. Formula DAX cálculo de Ranking top 10 para rating.....	52
Ilustración 52. Formula DAX cálculo de Ranking top 10 para share	52
Ilustración 53. Variables calculadas a nivel de tabla a través de las medidas....	53
Ilustración 54. Relación de filtros creados en el dashboard.....	54
Ilustración 55. Totales de contexto general.	55
Ilustración 56. Ranking del Top 10 de Rating Máximo	55
Ilustration 57. Ranking del top 10 del Share.	56
Ilustración 58. Comparativa indicadores por periodo.	56
Ilustración 59. Comparativo de totales generales mes a mes.	57
Ilustración 60. Matriz periodo , programas vs indicadores.....	57
Ilustración 61. Indicadores por programa.	58
Ilustración 62. Gráfico de los indicadores por franjas horarias.	59
Ilustración 63. Grafico de indicadores por franja horaria con filtros aplicados ...	59
Ilustración 64. Reporte detallado distribución indicadores y por periodo y franja horaria.	59
Ilustración 65. Página 1	60
Ilustración 66. Página 2	60
Ilustración 67. Pagina 3	61
Ilustración 68. Arquitectura implementada.....	61

Índice de tablas

Tabla 1. Definición de franjas horarias	7
Tabla 2. Comparativo propuesta dashboard vs empresas con soluciones similares en el mercado.	15

Resumen

Mayha Comunicaciones (El nombre de la empresa fue sustituido para mantener la confidencialidad y los datos usados para este trabajo no corresponde a la realidad del negocio) comercializa un producto de televisión por cable en la ciudad de Villavicencio, Colombia, y ha comenzado a trabajar con 97 empresas de programación que generan 163 canales que transmiten 1.440 minutos de programación diaria.

Lo que dificulta obtener métricas de la medición cuantitativa de la audiencia de un programa y la clasificación de los canales de televisión en función de su audiencia o desempeño.

La empresa está obligada a dar información sobre el seguimiento de los usuarios y la medición de la experiencia del cliente. Los datos son demasiado grandes, no se aplican las reglas comerciales y los procesos de cálculo se realizan trimestralmente, lo que dificulta que Mayha calcule estas métricas con precisión. El objetivo del trabajo es desarrollar un tablero que permita la visualización de los indicadores de rating y ranking de los canales de televisión de forma automatizada, precisa y confiable a partir de los datos generados por la empresa Mayha. El diseño e implementación del tablero será parte del trabajo.

Para la realización de la metodología de trabajo se utilizará la metodología por fases identificada en Dama DMBOK.

Palabras clave: Dashboard, Rating, Ranking, Análisis de datos

Abstract

Mayha Comunicaciones (The name of the company was substituted to maintain confidentiality and the information used for this work does not correspond to the reality of the business) markets a cable television product in the city of Villavicencio, Colombia, and has begun to work with 97 programming companies that generate 163 channels that broadcast 1,440 minutes of daily programming.

This makes it difficult to obtain quantitative measurement metrics of the audience of a program and the classification of television channels based on their audience or performance.

The company is required to provide information on user tracking and customer experience measurement. The data is too large, business rules are not applied, and calculation processes are done on a quarterly basis, making it difficult for Mayha to calculate these metrics accurately. The objective of the work is to develop a dashboard that allows the visualization of the rating and ranking indicators of the television channels in an automated, precise and reliable way from the data generated by the Mayha company. The design and implementation of the dashboard will be part of the job.

To carry out the work methodology, the phased methodology identified in Dama DMBOK will be used.

Keywords: Dashboard, Rating, Ranking, Data analysis

1 Introducción

El servicio de televisión por cable en Villavicencio es un nuevo desafío para Mayha Comunicaciones. Para brindar este servicio, Mayha trabaja en conjunto con 97 empresas de programación de televisión que generan 163 canales y emiten un total de 1.440 minutos diarios de programación continua.

Una gran cantidad de datos no está siendo procesada adecuadamente por la complejidad del proceso de actualización de la programación, sumado a la necesidad de cumplir con la entrega de datos normativos sobre monitoreo de usuarios y medición de la experiencia del cliente.

La falta de procesos de limpieza y clasificación de datos ha provocado que los indicadores de rating y ranking de los canales de televisión se calculan manualmente con altos márgenes de error en hojas de cálculo, lo que genera incertidumbre en los resultados.

El tablero que se desarrolló permite la visualización precisa de los indicadores de rating y ranking de los canales de televisión de Villavicencio, a partir de la información generada por Mayha Comunicaciones.

Se identificarán las fuentes de datos, se realizarán los procesos, se diseñarán las funciones y se implementará la herramienta para lograr este objetivo.

2 Descripción y formulación del problema

Actualmente Mayha Comunicaciones presta un servicio de televisión por cable en la ciudad de Villavicencio que lo ha llevado a asumir un nuevo reto, ya que la compañía está vendiendo un producto que antes no ofrecía en Colombia. Por esta razón Mayha Comunicaciones viene trabajando mancomunadamente con 97 programadoras de televisión que generan 163 canales que a su vez emiten un total de 1440 minutos diarios de programación continua.

Dentro de los acuerdos comerciales con algunos de los canales se comenzó a recibir la programación por diferentes medios; principalmente estos datos se pueden consultar por endpoint / web services (Richardson, 2017) lo que causa complejidad en el proceso ya que la frecuencia de actualización es variable: diaria, semanal o máximo quincenal en el caso de las programadoras institucionales con programación de baja rotación o cambio. Este primer conjunto de datos toma el nombre de EPG o (Electronic Programa Guide), el cual se almacena a diario y se conserva históricamente.

Por otra parte, Mayha como todas las empresas de TV por cable en Colombia está obligada a entregar información regulatoria con respecto al monitoreo del usuario, los contenidos que consume y la medición de la experiencia del cliente. Estos datos son únicamente confiables a través del cálculo del ranking y del rating. Por lo tanto, es su deber almacenar toda la interacción del televidente minuto a minuto con los dispositivos de TV como el decodificador y el control remoto, por ejemplo: el encendido del TV, el cambio de canales, el salto a la visualización de la guía, contabilización del tiempo por canal, visualización de imágenes de metadata, thrillers, acceso a programas especiales tipo “pague por ver”, compra de acceso a canales premium, procesos de grabación para disfrutar con posterioridad, etc. Toda esta interacción se almacena diariamente y hace parte de un segundo conjunto de base de datos.

La problemática principal de Mayha es que a pesar de tener todos estos conjuntos de datos mencionados anteriormente más los procesos de extracción desde los webservices y almacenamiento masivo, no es fácil el cálculo de los indicadores porque los datos son muy grandes, se encuentran sin procesos de limpieza, también, no se aplican las reglas de negocio que permitan segmentar y clasificar la información adecuadamente. Actualmente todos estos indicadores se calculan con muestras que se procesan manualmente con márgenes de error altos en hojas de cálculo de Excel, además, se procesan trimestralmente y no mensual como es requerido, causando incertidumbre en los resultados.

2.1 Formulación del problema

Cómo determinar el rating y ranking de los canales de televisión de manera precisa, confiable y automatizada en Villavicencio, a partir de la información generada por la empresa Mayha.

2.1.1 Objetivo General

Desarrollar un dashboard para la visualización de indicadores del rating y ranking de los canales de televisión en Villavicencio, a partir de los datos generados por la empresa Mayha.

2.1.2 Objetivos Específicos

- Identificar las fuentes de datos relevantes para el cálculo del rating y ranking de los canales de televisión en Villavicencio, a partir de los datos obtenidos de la empresa Mayha.
- Realizar los procesos de ETL que permitan obtener una data más consistente para el proceso de análisis sobre la información referente a los canales de televisión en Villavicencio.
- Diseñar las características funcionales y no funcionales del dashboard, a partir de la data consistente.
- Implementar el dashboard donde se visualiza el rating y ranking de los canales de televisión en Villavicencio.

3 Marco referencial

En el ámbito de la televisión, es fundamental contar con herramientas que permitan evaluar el rendimiento de los canales y su posición en el mercado. En este sentido, el desarrollo de un dashboard que visualice los indicadores del rating y ranking de los canales de televisión en Villavicencio se convierte en un objetivo clave para la empresa Mayha. Sin embargo, la construcción de este dashboard implica la consideración de diversos conceptos y teorías relacionados con la gestión de datos y el análisis de la industria televisiva. En este marco referencial, se abordarán dichos conceptos y teorías para contextualizar y fundamentar la solución propuesta al problema planteado.

3.1 Marco teórico

El problema que se presenta en este caso es desarrollar un dashboard para la visualización de indicadores del rating y ranking de los canales de televisión en Villavicencio, a partir de los datos generados por la empresa Mayha. Para poder abordar este problema, es necesario tener en cuenta varios conceptos clave relacionados con los canales de televisión y la gestión de datos.

En primer lugar, es importante entender qué son los canales de televisión y cómo funcionan: "Los canales de televisión son los medios de comunicación que transmiten programas de televisión a través de ondas electromagnéticas o señales de cable para que puedan ser vistos por los espectadores" (Gómez, 2017). Los canales de televisión generan una gran cantidad de datos relacionados con la audiencia, como el rating y el ranking, que son esenciales para el análisis, la toma de decisiones y el reporte a entidades de control como el Ministerio de Tecnologías de la Información y las Comunicaciones (MINTIC) (Martínez, 2019).

Para poder gestionar y analizar estos datos, es necesario utilizar herramientas y técnicas específicas: Primero, la identificación de fuentes de datos que es el proceso de recopilación de información sobre los programas de televisión y su audiencia para su posterior análisis. Segundo, Los procesos de ETL que se usan para extraer, transformar y cargar datos desde diferentes fuentes a una base de datos centralizada (García, 2020). Tercero, el procesamiento de estos conjuntos de datos en un gestor para que se pueden acceder fácilmente (Ramírez, 2018). Estos gestores de bases de datos son herramientas informáticas que se utilizan para crear, mantener y administrar bases de datos (Pérez, 2019).

Para poder analizar los datos de manera efectiva, es importante tener en cuenta la calidad de los datos. Que se refiere a la medida en que los datos son precisos, completos, coherentes y relevantes para el análisis (Gutiérrez, 2017). Para mejorar la calidad de los datos, es necesario realizar procesos de limpieza de datos, que consisten en identificar y corregir errores e inconsistencias en los datos (Hernández, 2020). Una vez que se tienen datos de calidad, es posible comenzar a realizar procesos de análisis de datos para determinar los indicadores de rating y

ranking de los contenidos que consumen los habitantes de Villavicencio esto se puede realizar a través de herramientas de visualización que son útiles para presentar los resultados de manera clara y fácil de interpretar (Sánchez, 2019).

En resumen, para abordar el problema planteado de determinar el rating y ranking de los canales de televisión de manera precisa, confiable y automatizada en Villavicencio, es necesario tener en cuenta varios conceptos clave relacionados con los canales de televisión y la gestión de datos como: Recopilar y gestionar datos de calidad utilizando herramientas y técnicas específicas, para luego realizar procesos de análisis de datos y presentar los resultados de manera clara y fácilmente interpretable. En el capítulo de Aspecto Metodológicos se dará una descripción más completa y detallada sobre todo el flujo de datos.

Teniendo en cuenta lo anterior se debe aclarar de manera específica los siguientes conceptos.

3.1.1 Canales de televisión:

Según la Real Academia Española, los canales de televisión son "cada una de las señales de televisión que transmiten programas a través de una frecuencia determinada" (Real Academia Española, s.f.). En este trabajo se puntualiza que Mayha Comunicaciones trabaja únicamente con conexiones de fibra óptica y dispositivos compatibles a este medio de transmisión. Por esta razón Mayha es una empresa que ofrece servicios de TV por cable (Rodríguez, 2021).

3.1.2 Decodificador:

Es un dispositivo que se instala en una ubicación física, conectada dentro de una red de fibra o tendido ópticos, y su principal función es procesar la información óptica y decodificar la señal hacia el televisor. Los decodificadores instalados por Mayha utilizan un microprocesador para ejecutar el software que controla las funciones del decodificador y la interfaz de usuario. Actualmente, la empresa instala un dispositivo de marca Huawei referencia, tv ont Gpon Huawei hg8245h. (Pérez, 2018; Huawei, 2023).

3.1.3 Rating:

Es una medida de la audiencia que indica el número de personas que están sintonizando un programa de televisión en un momento determinado. Por lo general, se expresa en porcentaje. Existen varias formas de medir el rating:

- **Share:** También se le conoce como "cuota de pantalla" permite tomar la medida de cuántos dispositivos están sintonizados en un canal y programa a una hora específica sobre todos los dispositivos que conforman el total de clientes que pueden sintonizar el mismo canal y que tienen servicio activo (Ruiz, 2016).
- **Índice de audiencia:** según el sitio web de la Asociación para la Investigación de Medios de Comunicación (AIMC), el índice de audiencia es el porcentaje de

la población que está viendo un determinado programa en un momento determinado. Se calcula dividiendo el número de personas que están viendo el programa por el número total de personas que tienen la posibilidad de verlo (Dircomfidencial, 2020). Para este índice Mayha utiliza la información de planta ya que es la fuente de datos de dispositivos completa que permite determinar el estado de los servicios, en este caso activos e inactivos.

Para Mayha Comunicaciones es importante tomar la medida de los dos indicadores.

3.1.4 Ranking:

Es una clasificación de los canales de televisión según su popularidad o audiencia. Se establece a partir del rating. Lo que se logra con el ranking es organizar en un top máximo o mínimo las diferentes mediciones del rating (Flores, 2015, pág. 103).

3.1.5 Addons:

Los addons son los canales y programas que se pagan como suscripciones adicionales por ej: HBO, Win Sport, Paramount etc. Generalmente en la medición de rating y ranking debe realizarse por separado para evitar cálculos incorrectos ya que no todos los usuarios acceden a estos productos.

3.1.6 Recopilación de orígenes de datos:

Es el proceso de recopilación de información sobre los programas de televisión y su audiencia para su posterior análisis. Es necesario para la medición de indicadores de audiencia contar con 3 orígenes básicos de datos, necesarios para medir los indicadores (Cavaller, 2014):

- **EPG (Electronic Program Guide):** Es la programación por canal y por fecha y hora, almacenada en una base de datos, es de carácter volátil ya que los programas pueden tener cambios abruptos: como cambios de programación de último momento, noticias de última hora, cambios de frecuencia o de hora según las estrategias que tomen las programadoras de TV. Por este motivo la EPG se mantiene históricamente en la base de datos y de esta manera se logra determinar cuál fue el canal/programa/horario que efectivamente se emitió.
- **Event:** Es el nombre que se da al conjunto de datos que hace referencia a los datos de eventos desencadenados por el televidente en los dispositivos instalados, por ejemplo: decodificador y control remoto. Esta información se almacena y conforma grandes volúmenes de datos; los eventos desencadenados por los usuarios consisten en todas las interacciones minuto a minuto y se registra la cantidad de visitas del televidente por canal, franja y programa.

- **Devices:** Es la base de datos que contiene la información puntual del servicio, las llaves del dispositivo con las que identificar el movimiento del usuario en la base de **Event**, así como el estado del servicio y disponibilidad del dispositivo ya que pueden encontrarse servicios activos y suspendidos. Con estos estados definen el total de servicios disponibles a la hora de hacer la medición del rating.

3.1.7 Franjas Horarias:

Los gustos y consumos de los clientes cambian según el contenido de los canales y el tiempo disponible para acceder a ellos, como, por ejemplo: las franjas horarias utilizadas para ver canales de TV. Las franjas horarias que están parametrizadas para Colombia (Banrep,2013) están clasificadas en la siguiente tabla (ver tabla 1).

Tabla 1. Definición de franjas horarias

<p align="center">FRANJAS HORARIAS LUNES - VIERNES</p>	<p>DAY: 06 AM -11:59 AM EARLY: 12 M - 18:59 PM PRIME: 19 PM - 22:59 PM LATE : 23:00 PM - 23:59 PM OVERNIGHT: 00:00 AM - 05:59 AM</p>
<p align="center">FRANJAS HORARIAS FINES DE SEMANA</p>	<p>DAY: 06 AM -11:59 AM EARLY: 12 M - 17:59 PM PRIME: 18 PM - 21:59 PM LATE : 22:00 PM - 23:59 PM OVERNIGHT: 00:00 AM - 05:59 AM</p>

Fuente: Elaboración propia

3.1.8 Bases de datos:

Es una colección de datos interrelacionados entre sí, estructurados de manera particular que cumple con lo que se conoce ampliamente como normalización, con la cual puede representarse un modelo lógico para que la información pueda ser recuperada fácilmente a través de consultas (Gómez, 2013). Cumpliendo con estos lineamientos generales en esta definición son varias las empresas del sector informático como: Microsoft, Oracle, PostgreSQL entre otros, que han desarrollado herramientas denominadas sistemas de administración de bases de datos (SGBD) (Gómez, 2013). Mayha utiliza este tipo de tecnologías y mantiene almacenada en bases de datos y las gestiona a través de los SGBD.

3.1.9 Sistemas de administración de bases de datos (SGBD):

Los gestores de bases de datos son herramientas informáticas que se utilizan para crear, mantener y administrar bases de datos. (Gómez, 2013). En el caso del problema planteado, se utilizará para almacenar la información de los programas de televisión, los eventos del usuario, y los datos de los suscriptores en el motor

SQL Server 2019. Esta herramienta fue desarrollada por Microsoft y se encuentra posicionada dentro de los líderes en motores de bases de datos y se encuentra presente en la mayoría de las organizaciones de los sectores más prestantes del mercado.

3.1.10 Modelo de datos:

Es una representación gráfica de la estructura de los datos, que permiten describir la composición de la información a un alto nivel para que sea más comprensible para las personas; cuando se hace referencia a modelo de datos, se hace referencia a la lógica matemática con la que se describe la composición de los datos dentro del modelo para que sea comprensible para el ser humano. Por otra parte, si tenemos un modelo a alto nivel, también existe una representación de esta misma data a bajo nivel que detalla cómo se almacena a nivel de máquina. (Gómez, 2013).

En Mayha existen diversos tipos de modelo de datos, desde estructurados, semi estructurados y no estructurados; para este trabajo que describe el cálculo del índice de audiencia y clasificación para la medición de la audiencia, se ha definido el modelo de datos relacional. Para asegurar la integridad de los datos y evitar problemas de inserción, actualización o eliminación, se aplica el proceso de normalización hasta la tercera forma normal.

Existen diferentes niveles de formas normales, siendo las más conocidas la primera forma normal (1FN), la segunda forma normal (2FN) y la tercera forma normal (3FN) (Escobar et al., 2012).

La primera forma normal (1FN): Establece que cada atributo de una tabla debe contener un único valor. Esto significa que las celdas de la tabla no deben contener valores repetidos o listas de valores. Si una tabla viola esta regla, se puede dividir en varias tablas para cumplir con la 1FN.

La segunda forma normal (2FN): Establece que una tabla debe tener una clave primaria única y que cada atributo de la tabla debe depender funcionalmente de la clave primaria, es decir, que cada atributo de la tabla debe estar relacionado directamente con la clave primaria.

La tercera forma normal (3FN): Establece que no debe haber dependencias funcionales transitivas entre los atributos, es decir, que no debe haber una relación indirecta entre dos atributos a través de un tercero.

Existen también formas normales más avanzadas, como la cuarta forma normal, la quinta forma normal (5FN) y la forma normal de Boyce-Codd (BCNF), que establecen requisitos más estrictos para el diseño de la base de datos y permiten una mayor reducción de la redundancia de datos y la mejora de la consistencia de los datos. En general, seguir las formas normales en el diseño de una base de datos es una buena práctica para asegurar la integridad y eficiencia de la información almacenada en ella. (Gómez, 2013).

3.1.11 Proceso de Extract, Transform, Load (ETL):

ETL se utiliza en la construcción y mantenimiento de los datos. ETL es un acrónimo que significa Extraer, Transformar y Cargar.

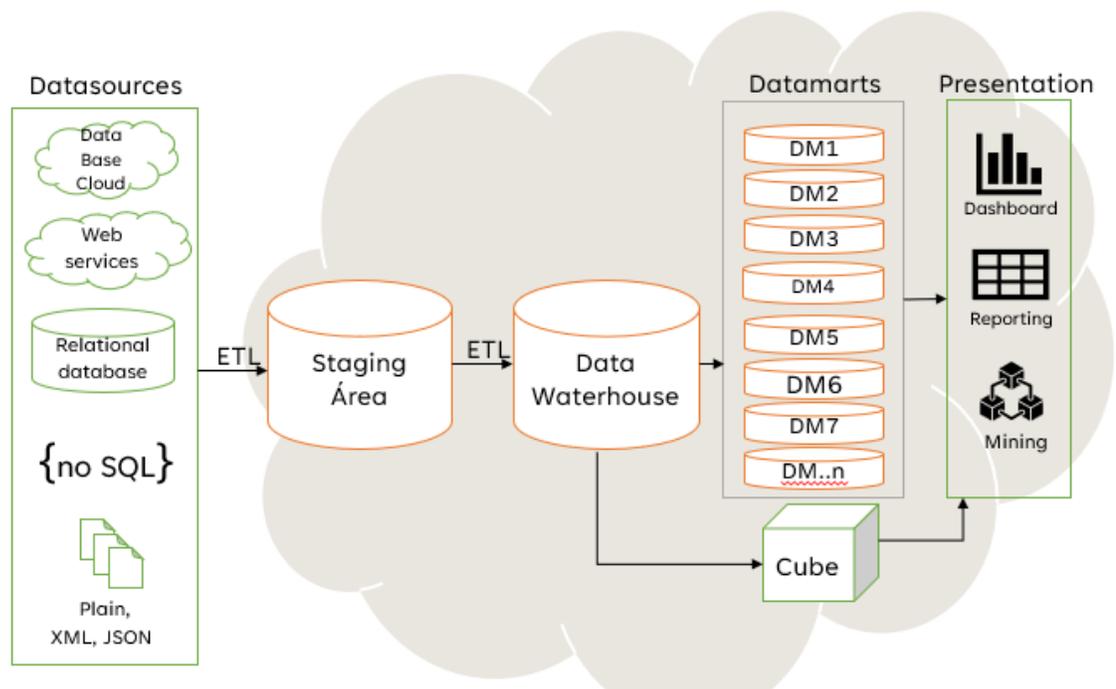
En este proceso, los datos son extraídos de diversas fuentes de datos, tales como bases de datos operativas, archivos, aplicaciones, etc. Una vez que se han extraído los datos, se aplican transformaciones para asegurar que los datos son coherentes y están en el formato adecuado para su posterior análisis. Por ejemplo, pueden realizarse transformaciones como la limpieza de datos, la eliminación de duplicados, la agregación de datos, etc. Este proceso es fundamental para garantizar que los datos almacenados sean precisos, consistentes y estén en el formato adecuado para su análisis posterior. (Curto, 2012).

En el presente trabajo de grado, se hace referencia a procesos ETL ya que los datos obtenidos de las programadoras son procesados para aplicar transformaciones orientadas a la limpieza y clasificación de datos.

El proceso siguiente (Ver Ilustración 1) describe a grandes rasgos el contexto en el que funciona los procesos ETL, las flechas negras indican de izquierda a derecha como estos datos se trasladan desde diversas fuentes a otras bases de datos o estructuras especiales y que finalmente se entregan para análisis del usuario.

Mayha no es ajena a este tipo de tecnologías las cuales están presentes en su arquitectura.

Ilustración 1. Ejemplo de arquitectura de datos corporativa tradicional.



Fuente: Elaboración propia.

3.1.12 Herramientas de ETL:

Son programas informáticos que se utilizan para automatizar el proceso de ETL. Las herramientas ETL son ideales para procesar grandes cantidades de datos, como bases de datos, archivos y sistemas de información. Estas herramientas proporcionan una forma eficiente de mover los datos entre sistemas. (Curto, 2012). En el mercado se pueden encontrar una gran diversidad de herramientas y con diferentes tipos de licenciamiento, entre las más populares encontramos: Integration Services (SSIS), ETL Talend Data Integration, Integrate.IO entre otras. Para la extracción de los datos requeridos para el cálculo de rating y ranking en este trabajo se utilizó Integration Services (SSIS).

3.1.13 Calidad de datos:

La calidad de los datos se refiere a la exactitud, integridad, consistencia y relevancia de los datos que se utilizan para realizar análisis y tomar decisiones. El tema de calidad de datos en su completitud puede componerse de múltiples variables y técnicas desde las relacionadas con el gobierno de datos, cultura organizacional entre otras. En este apartado solo se enfocará al proceso técnico

puntual de limpieza, clasificación y segmentación de la información que se programa desde la herramienta ETL.

Estas herramientas son útiles para mejorar la calidad de la información, ya que permiten a los usuarios filtrar y fusionar datos de diferentes fuentes para asegurarse de mejorar la precisión y consistencia. Las herramientas ETL también se pueden utilizar para detectar y corregir errores, eliminar datos duplicados y aplicar reglas de negocio y reducir márgenes de error. (Curto, 2012).

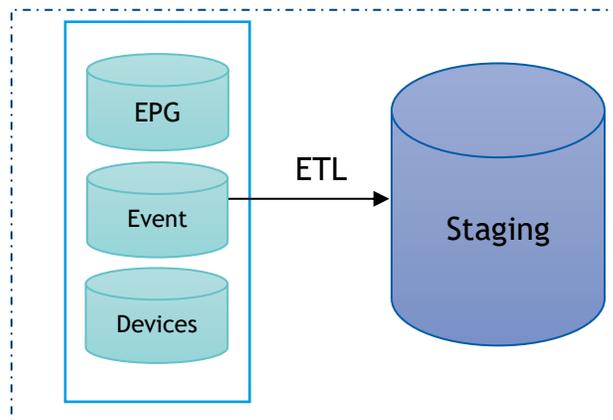
Los procesos de calidad a través de las herramientas ETL que se aplicaran para el procesamiento de datos que son insumos para el cálculo del rating y ranking básicamente tienen que ver con:

1. **Procesamiento de registros con información en NULL**, reemplazo o marcado de este tipo de valor para darle mayor significado a la data.
2. **Procesamiento de registros duplicados**, Para registros idénticos se realiza el retiro de uno de los registros duplicados.
3. **Registros sin contenido**, Para registros en base de datos que no tienen ningún tipo de información se procede a retirarlo ya que no aporta ningún significado valioso al análisis.
4. **Verificación de conexión entre insumos**, se hace un cruce de la información entre las fuentes de datos para verificar su correlación; según los hallazgos la data no correlacionada y que no aporte ningún significado al análisis se retira.

Más adelante se dará detalle de todo este proceso junto con la aplicación de reglas de negocio para Mayha Comunicaciones, hasta aquí en este apartado se explica el proceso de cómo la herramienta de ETL contribuirá en primera instancia con el proceso de calidad de datos.

Como se describe a continuación (Ver Ilustración 2). Dando una aproximación del proceso de cálculo de rating para Mayha, se procederá con la extracción de las fuentes de programación, eventos y dispositivos y se aplicarán las reglas descritas arriba para obtener una versión de datos limpia y se almacenará en una base denominada Staging Area o base de datos de preparación. Sobre esta se realizará la aplicación de reglas de negocio.

Ilustración 2. ETL reglas de limpieza hacia Staging



Fuente: Elaboración propia.

3.1.14 Limpieza de datos:

Es el proceso de identificar y corregir errores e inconsistencias en los datos. La limpieza de datos es importante para garantizar la calidad de los datos. (Curto, 2012).

3.1.15 Base de datos Staging:

Las bases de datos de Staging se usan para almacenar datos desde una variedad de fuentes. Estos datos se limpian, transforman y validan antes de ser cargados en la base de datos de Staging. Esta limpieza, transformación y validación pueden incluir la corrección de errores, la eliminación de datos redundantes, la combinación de datos de diferentes fuentes, etc.

La ventaja de una base de datos de Staging es que puede proporcionar datos limpios y estructurados para informes y análisis. Esto significa que los informes y análisis serán más precisos y confiables. Esta base de datos también puede ayudar a reducir los costos de almacenamiento de datos, ya que los datos se almacenan en un solo lugar y no se necesita almacenar los datos en diferentes sistemas. (Curto, 2012).

3.1.16 Herramientas de visualización:

Son programas informáticos que se utilizan para presentar los datos de forma gráfica y visualmente atractiva. Las herramientas de visualización son útiles para presentar los resultados del análisis de datos de manera clara y fácilmente interpretable. (Sandell, 2020). Puntualmente para este trabajo se utilizó Power BI.

3.2 Estado del arte

El rating y ranking de canales de televisión en Colombia ha sido un tema de gran importancia en la industria de la televisión en los últimos años. A través de estos indicadores, se puede conocer el nivel de audiencia que tienen los canales y programas de televisión, lo que se traduce en una herramienta clave para la toma de decisiones de los anunciantes y programadores.

En Colombia, existen diferentes empresas encargadas de medir el rating y ranking de la televisión, entre las que se destacan Kantar IBOPE Media, Nielsen y ComScore. Estas empresas utilizan diferentes técnicas y metodologías para recopilar y analizar los datos, incluyendo encuestas, mediciones de audiencia en tiempo real y análisis de datos a partir de registros de transacciones.

Además, la industria de la televisión ha evolucionado con el tiempo, lo que ha llevado a cambios en la forma en que se mide y se utiliza el rating y ranking. Por ejemplo, con la aparición de nuevas tecnologías como el streaming y la televisión por internet, se ha vuelto cada vez más importante medir el consumo de televisión en diferentes plataformas y dispositivos, lo que ha llevado a la necesidad de desarrollar nuevas técnicas y herramientas para la medición de audiencia.

En cuanto a la aplicación del rating y ranking en Colombia, estos indicadores son utilizados por diferentes actores de la industria de la televisión, incluyendo canales de televisión, anunciantes y agencias de publicidad. La información obtenida a partir del rating y ranking es utilizada para tomar decisiones sobre la programación de los canales, la selección de programas y la inversión en publicidad.

En conclusión, el rating y ranking de canales de televisión en Colombia es un tema de gran importancia para la industria de la televisión y ha evolucionado con el tiempo a medida que han surgido nuevas tecnologías y cambios en la forma en que se consume la televisión. Diferentes empresas se encargan de medir y analizar estos indicadores, y su aplicación se extiende a diferentes actores de la industria de la televisión.

En Colombia existen varias empresas que se encargan de medir el rating y ranking de los canales de televisión, entre las más reconocidas se encuentran:

3.2.1 Kantar IBOPE Media:

Esta empresa es una de las líderes en medición de audiencia a nivel mundial. En Colombia, se encarga de medir el rating y ranking de la televisión a través de su sistema "PeopleMeter", que recopila datos de audiencia de forma continua y en tiempo real, por medio de un dispositivo electrónico que se instala en los hogares seleccionados para formar parte de la muestra. El dispositivo registra automáticamente la sintonización de los canales de televisión y envía esta información a Kantar IBOPE Media, donde se procesa y analiza para generar los

ratings y rankings (Kantar IBOPE Media Colombia. (s.f.). Audiencia de TV. Kantar IBOPE Media Colombia).

3.2.2Nielsen:

Otra de las empresas líderes en medición de audiencia a nivel mundial. En Colombia, se encarga de medir el rating y ranking de la televisión a través de su sistema de medición de audiencia, que utiliza una muestra de hogares representativos para medir la audiencia de televisión. La muestra se selecciona cuidadosamente para reflejar la diversidad de la población y se instala un dispositivo electrónico en cada hogar para registrar la sintonización de los canales de televisión. Los datos recopilados se procesan y analizan para generar los ratings y rankings (Nielsen. (s.f.). National TV. Nielsen.).

3.2.2ComScore:

Esta empresa se encarga de medir la audiencia en diferentes plataformas, incluyendo televisión, internet y dispositivos móviles. En Colombia, se encarga de medir el rating y ranking de la televisión a través de su sistema de medición de audiencia.

Esta empresa utiliza un panel de usuarios para medir la audiencia de televisión. Los usuarios seleccionados instalan un software especial en sus dispositivos para permitir la medición de la audiencia. ComScore combina los datos de su panel con otros datos de fuentes múltiples, como los operadores de cable, para generar los ratings y rankings (Comscore. (s.f.). Ranking de medios auditados por Comscore en Colombia).

3.2.3Tabla de resumen comparativo propuesta para Mayha vs empresas en el mercado:

El análisis comparativo se realizó teniendo en cuenta las diferencias y semejanzas entre la propuesta de desarrollo del dashboard para Mayha Comunicaciones y las soluciones existentes en el mercado para la medición de los indicadores de rating y ranking de canales y programas de televisión. Las empresas relacionadas en el presente estado del arte son las más representativas a nivel nacional e internacional.

Las semejanzas y diferencias se clasifican según las herramientas tanto de software y hardware adicionales que se deben instalar, el nivel de detalle en la segmentación poblacional por dispositivo y la disponibilidad en términos de propiedad y autoría de las herramientas como se representa a continuación. (Ver Tabla 2).

Tabla 2. Comparativa propuesta dashboard vs empresas con soluciones similares en el mercado.

Aspectos relevantes de las herramientas para la medición	Cantar IBOPE Media	Nielsen	ComScore	Dashboard Mayha
Instalación software o hardware adicional a los decodificadores de Mayha.	SI	SI	SI	NO
Segmentación de población: Conteo de miembros del hogar por dispositivo instalado.	SI	SI	SI	NO
Otorga autoría y propiedad software para el operador por cable Mayha.	NO	NO	NO	SI

Fuente: Elaboración propia

3.3 Impacto

Mayha a pesar de capturar la información y almacenarla dentro de la plataforma que posee no tiene automatizado el proceso de cálculo de rating y ranking, por lo tanto, la automatización utilizando consultas de base de datos y procesamiento automático mejora significativamente la precisión, la eficiencia, la escalabilidad, la personalización y la toma de decisiones. Esto conduce a una mayor productividad y competitividad para la empresa al adoptar esta tecnología. Sin embargo, el mayor impacto tendrá que ver con la capacidad de mantener actualizados los indicadores en periodos aceptables de tiempo, también, la de apreciar la tendencia y los niveles de éxito que tienen los canales y sus programas en los diversos horarios.

3.4 Componente de Innovación

En Mayha Comunicaciones, la innovación radica en la implementación de la automatización en el proceso de almacenamiento de la data, utilizando modelos especializados para análisis y la ejecución de algoritmos. Esto reemplaza las tareas manuales que antes eran realizadas por los funcionarios, quienes utilizaban herramientas inadecuadas como hojas de cálculo y procesadores de texto. Estas herramientas presentaban problemas de precisión y calidad de los datos, además de una frecuencia bimestral que dificultaba la comprensión de la información.

La implementación de este enfoque automatizado permite a la compañía utilizar la información de manera más efectiva en la toma de decisiones a nivel gerencial y en la gestión de contratos con los canales de televisión. Esto se traduce en beneficios económicos para la organización, ya que se evita incurrir en los altos costos asociados con el proceso manual utilizado por otras compañías. En resumen, se logra un ahorro económico significativo mientras se mejora la calidad y precisión de los datos utilizados en el proceso.

4 Metodología

Para el desarrollo del dashboard de medición de ranking y rating, se utilizó la metodología por fases identificada en el Dama DMBOK (DAMA, 2022). A continuación, se describen cada una de las fases.

4.1 Identificación de datos

Esta etapa se centra en la recopilación y selección de datos relevantes. En ella se identifican y catalogan los datos existentes en una organización para comprender mejor los datos que se utilizan y cómo se relacionan. Es un paso crucial en la gestión de datos que permite tomar decisiones informadas sobre cómo administrar y utilizar los datos de manera efectiva.

4.2 Consulta de los datos

Esta etapa consiste en el proceso de buscar y extraer información específica de una base de datos mediante la ejecución de fórmulas y funciones al sistema de gestión de bases de datos. Es fundamental para el análisis y cálculo de indicadores de la compañía.

4.3 Alistamiento de los datos

En este proceso se ejecuta la preparación de los datos para su análisis, que incluye tareas como la limpieza, transformación y organización de los datos. El objetivo es asegurar que los datos sean precisos, completos y coherentes para el análisis. Es una parte esencial del proceso de análisis de datos.

4.4 Almacenamiento

Esta etapa es una parte esencial del ciclo de vida de los datos, ya que permite salvar la información de manera permanente dentro los dispositivos dedicados para este fin, como discos duros, memorias etc. Este proceso está relacionado con la creación, persistencia, modificación y destrucción de la misma.

4.5 Procesamiento

Esta etapa se refiere al conjunto de operaciones que se realizan sobre los datos para transformarlos y analizarlos con el fin de obtener información útil. Esto puede incluir operaciones como la agregación, el filtrado, la clasificación.

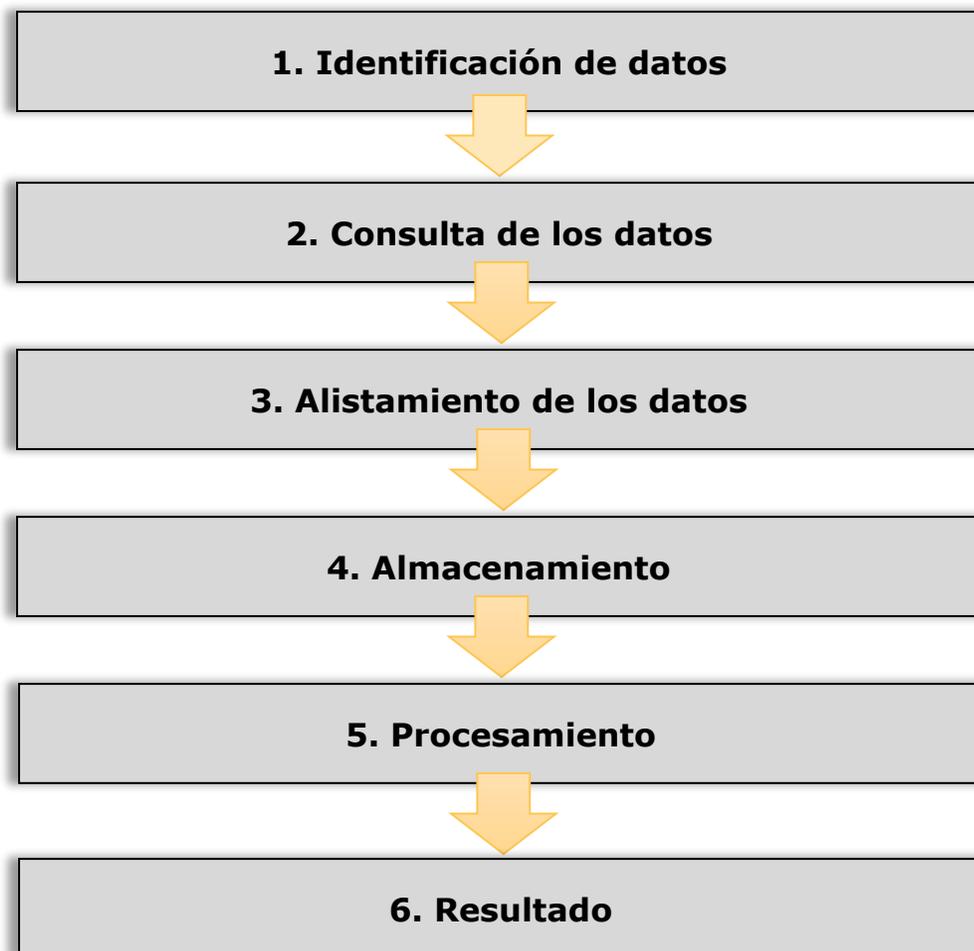
4.6 Resultado

Esta etapa se refiere al producto final del análisis de datos, que proporciona información útil y valiosa para la toma de decisiones en una organización. El

resultado puede ser un informe, un gráfico, un cuadro de mando o cualquier otro tipo de visualización de datos que muestre los hallazgos del análisis.

A continuación, se describe gráficamente (Ver Ilustración 3) las fases de la metodología a utilizar.

Ilustración 3. Fases ejecutadas para el proyecto.



Fuente: Elaboración Propia

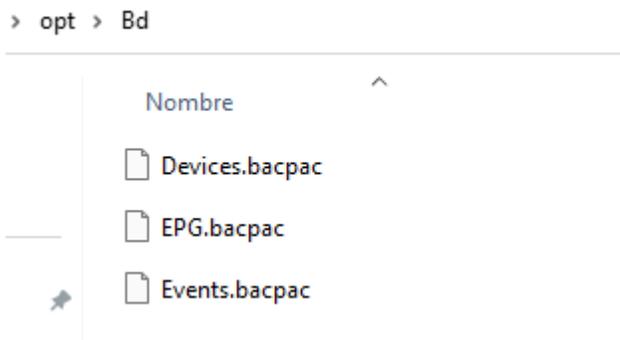
5 Desarrollo de la propuesta

En la propuesta de desarrollo del dashboard de medición de ranking y rating se ha elegido seguir la metodología por fases que se encuentra establecida en el Dama DMBOK (DAMA, 2022). La elección de esta metodología se debe a que permite una gestión adecuada de los recursos, asegurando la calidad y eficiencia del proyecto. A continuación, se describen cada una de las fases desarrolladas en la solución del problema planteado.

5.1 Identificación de Datos

La etapa de Identificación de Datos en la creación del Dashboard para Mayha, implica la recolección y selección de datos relevante para su posterior uso en la herramienta. Para llevar a cabo esta tarea, se seleccionaron tres fuentes de datos (EPG, EventChannel, Devices). En particular, se obtuvieron los datos a través de la empresa Mayha Comunicaciones (Ver Ilustración 4), quienes proporcionaron por medio de la vía de backups y scripts estos datos y se presentaron en un formato determinado el cual fue procesado para generar las bases a con la información identificada según las reglas de negocio descritas en función a las prácticas de gobierno de datos que se plasmaron en un documento (Ver Anexo 4).

Ilustración 4. Entrega de Datos realizada por Mayha



Fuente: Elaboración propia

A continuación, se detallará el lugar de procedencia de los datos y sus atributos distintivos.

5.2 EPG

Es una base de datos que contiene la programación de televisión organizada por canal, fecha y hora, y se mantiene históricamente con el fin de asegurar que la información tenga todos los cambios efectuados por las programadoras de TV en la dinámica de presentación y cambio súbito de la programación.

Los cambios de programación obedecen a estrategias de publicidad y mercadeo en la búsqueda de hacer que los televidentes sintonicen, vean los contenidos y que a su vez consuman el espacio dedicado a la publicidad. Esta es una de las razones por las cuales tener un correcto cálculo de los indicadores de rating rankeado permite explotar al máximo los ingresos por publicidad en Mayha. Primero lanzando estrategias junto con las programadoras y segundo revisando el éxito obtenido en las emisiones, dándole la posibilidad a los anunciantes de posicionar mejor sus productos y servicios realizando inversión en Mayha según las mediciones de audiencia (Ver ilustración 5).

Ilustración 5. Resultset de la fuente entregada por Mayha de EPG.

```
SELECT [IdCanal],
       [IdPrograma],
       [AnioInicioPrograma],
       [MesInicioPrograma],
       [DiaInicioPrograma],
       [HoraInicioPrograma],
       [MinutoInicioPrograma],
       [AnioFinPrograma],
       [MesFinPrograma],
       [DiaFinPrograma],
       [HoraFinPrograma],
       [MinutoFinPrograma],
       [DuracionPrograma]
FROM [IPTV].[EPGHistorico]
```

IdCanal	IdPrograma	AnioInicioPrograma	MesInicioPrograma	DiaInicioPrograma	HoraInicioPrograma	MinutoInicioPrograma
1805	509853	2022	11	8	21	0
1805	510279	2023	1	10	14	0
1805	517909	2022	10	30	16	30

Fuente: Elaboración propia

Para información sobre los metadatos de la tabla EPG se recomienda ver el diccionario de datos de esta tabla en el **Anexo 1**.

5.3 Event

Es el nombre dado por Mayha a los datos de eventos desencadenados por los televidentes en dispositivos como decodificadores a través del control remoto, almacenados en la base de datos. Es importante recalcar que la base **Event** entrega la información de decodificadores que sintonizaron canales desde al menos 1 segundo hasta los 5 minutos marcando este comportamiento como “**Surfing**” e indica la situación en el que el usuario está cambiando canales en búsqueda de algún programa de su interés. Y los programas con más de 5 minutos los clasifica como “**Tuning**” que significa sintonizado. Además, incluye en el campo **StripStar** la franja horaria. (Ver Ilustración 6).

Ilustración 6. Resultset TOP 10 de la fuente entregada por Mayha tabla events

	IdEvent	TypeDevice	MACDevice	SerialDevice	StartNameDay	StartTuningTime	EndTuningTime	EndNameDay	Duration	Type Tuning	StripStart
1	1	Set Top Box	44:43:93:94	29374	Saturday	2022-10-01 00:00:00.000	2022-10-01 00:08:00.000	Saturday	8	Tuning	OverNight
2	2	Set Top Box	41:40:06:89	11025	Saturday	2022-10-01 00:00:00.000	2022-10-01 00:09:00.000	Saturday	9	Tuning	OverNight
3	3	Set Top Box	89:77:86:72	33442	Saturday	2022-10-01 00:00:00.000	2022-10-01 00:09:00.000	Saturday	9	Tuning	OverNight
4	4	Set Top Box	18:64:87:84	25053	Saturday	2022-10-01 00:00:00.000	2022-10-01 00:11:00.000	Saturday	11	Tuning	OverNight
5	5	Set Top Box	42:57:17:11	80512	Saturday	2022-10-01 00:00:00.000	2022-10-01 00:14:00.000	Saturday	14	Tuning	OverNight
6	6	Set Top Box	81:35:83:66	99029	Saturday	2022-10-01 00:00:00.000	2022-10-01 00:15:00.000	Saturday	15	Tuning	OverNight
7	7	Set Top Box	94:26:20:53	44006	Saturday	2022-10-01 00:00:00.000	2022-10-01 00:20:00.000	Saturday	20	Tuning	OverNight
8	8	Set Top Box	37:07:38:29	91738	Saturday	2022-10-01 00:00:00.000	2022-10-01 00:24:00.000	Saturday	24	Tuning	OverNight
9	9	Set Top Box	11:39:94:85	80328	Saturday	2022-10-01 00:00:00.000	2022-10-01 00:24:00.000	Saturday	24	Tuning	OverNight
10	10	Set Top Box	42:64:45:68	47440	Saturday	2022-10-01 00:00:00.000	2022-10-01 00:24:00.000	Saturday	24	Tuning	OverNight

Fuente: Elaboración propia.

Para información sobre los metadatos de la tabla Events se recomienda ver el diccionario de datos de esta tabla en el **Anexo 2**.

5.4 Devices

Es una tabla de la base de datos que almacena información puntual del dispositivo de tv denominado “**Set Top Box**” con el que se determina que se trata de un decodificador de TV, lo que permite definir el total de servicios disponibles para medir el rating (Ver Ilustración 7).

Ilustración 7. Resultset fuente entregada por Mayha denominada InventoryDevice.

```

SELECT [MACDevice]
      ,[SerialDevice]
      ,[TypeDevice]
      ,[CodCity]
      ,[Neighborhood]
      ,[EconomicCode]
FROM [Devices].[dbo].[InventoryDevice]

```

	MACDevice	SerialDevice	TypeDevice	CodCity	Neighborhood	EconomicCode
1	27:67:40:56	10006	Set Top Box	50001	El Camen	2
2	60:27:62:76	10008	Set Top Box	50001	El Camen	2
3	24:24:81:20	10009	Set Top Box	50001	El Camen	2
4	50:61:17:16	10018	Set Top Box	50001	El Camen	2
5	89:67:55:66	10020	Set Top Box	50001	El Camen	2

Fuente: Elaboración propia.

Para información sobre los metadatos de la tabla Devices se recomienda ver el diccionario de datos de esta tabla en el **Anexo 3**.

5.5 Consulta de los datos:

Durante la etapa de consulta de datos, se llevará a cabo la extracción de información desde el origen de los datos. Con el objetivo de manejar el gran volumen de data acumulativa, se utilizará la plataforma Azure de Microsoft para trabajar en la nube. Durante el proceso de lectura, se ejecutarán diversos procesos de almacenamiento temporal en la memoria del servidor, permitiendo lecturas a nivel lógico y físico para una gestión óptima de los datos.

En este punto se tendrá en cuenta el siguiente lineamiento estipulado en la política generada por el comité de gobierno de datos (Ver Anexo 5).

Para realizar y ejecutar el proceso descrito en la política en nube se deben realizar los siguientes pasos:

1) Crear un grupo de recursos:

En este caso el nombre de **ProyectoUan2023** se refiere al grupo de recursos en Azure utilizado para el proyecto en cuestión. Este nombre es una etiqueta de identificación para los recursos relacionados con el proyecto y permite una fácil gestión y organización de estos. Al utilizar esta nomenclatura, se sigue una convención consistente que ayuda a los usuarios a identificar rápidamente los recursos relacionados con el proyecto (Ver Ilustración 8).

Ilustración 8. Creación del grupo de recursos en Azure.

The screenshot shows the Azure portal interface for a resource group named 'ProjectoUan2023'. The left sidebar shows a list of resource groups with 'ProjectoUan2023' selected. The main content area displays the configuration for this resource group. Key details include:

- Información esencial:**
 - Suscripción: Azure for Students
 - Id. de suscripción: a8cd9e31-38d0-4562-8a36-5c17a1f01b6f
 - Ubicación: East US 2
 - Etiquetas: Proyecto: Ranking_Rating
- Recursos:**
 - psseprycan: SQL Server
 - Rating_EPG_Base_01 (psseprycan/Rating_EPG_Base_01): Base de datos SQL

Fuente: Elaboración propia.

2) Servidor SQL:

Se crea un servidor SQL bajo la nomenclatura "psseprycan", que se apalanca en la convención camel y su significado es fácilmente identificable como un servidor SQL productivo para el proyecto de canales, lo que facilita su gestión y control. Esta práctica se enmarca en la taxonomía de gobierno de datos al promover la consistencia y estandarización en la denominación de los recursos, lo que facilita su identificación y administración (Ver Ilustración 9).

Ilustración 9. Creación del servidor SQL

The screenshot shows the configuration page for the 'psseprycan' SQL Server. The left sidebar shows the navigation menu with 'Información general' selected. The main content area displays the configuration details:

- Información esencial:**
 - Grupo de recursos: ProyectoUan2023
 - Estado: Disponible
 - Ubicación: East US
 - Suscripción: Azure for Students
 - Id. de suscripción: a8cd9e31-38d0-4562-8a36-5c17a1f01b6f
 - Etiquetas: Haga clic aquí para agregar etiquetas.
- Administración del servidor:**
 - Administrador del servidor: Administrador
 - Redes: Mostrar configuración de redes
 - Administrador de Active Directory: No configurado
 - Nombre del servidor: psseprycan.database.windows.net

Fuente: Elaboración propia.

3) Base de datos Rating_EPG_Base_01:

Llamada "Rating_EPG_Base_01" para el proyecto se apalanca en la convención camel y su nombre es fácilmente identificable, ya que indica que se trata de una base de datos relacionada con el rating y la programación de la guía electrónica de programación (EPG). Esta práctica se enmarca en la taxonomía de gobierno de

datos ya que promueve la consistencia y estandarización en la denominación de los recursos, lo que facilita su identificación y administración, y ayuda a evitar confusiones y errores en la gestión de los datos (Ver Ilustración 10).

Ilustración 10. Creación de base de datos Rating_EPG_Base_01 en nube con marcado indicado en la política Nro. 2.

Inicio > ProyectoUan2023 > Marketplace > Azure SQL > Seleccione una opción de implementación de SQL >

Crear base de datos SQL

Microsoft

una cuenta gratuita de Azure? [Más información](#)

Detalles del proyecto

Seleccione la suscripción para administrar recursos implementados y los costes. Use los grupos de recursos como carpetas para organizar y administrar todos los recursos.

Suscripción *

Grupo de recursos * [Crear nuevo](#)

Detalles de la base de datos

Indique la configuración necesaria para esta base de datos, incluida la selección de un servidor lógico y la configuración de los recursos de proceso y almacenamiento.

Nombre de la base de datos *

Servidor * [Crear nuevo](#)

¿Quiere usar un grupo elástico de SQL? Sí No

Proceso y almacenamiento * **Básico**
Almacenamiento: 2 GB
[Configurar base de datos](#)

Redundancia del almacenamiento de copias de seguridad

Elija el modo de replicación de las copias de seguridad de PITR y LTR. La restauración geográfica o la posibilidad de recuperación tras una interrupción regional solo están disponibles si se ha seleccionado el almacenamiento con redundancia geográfica.

Redundancia de almacenamiento de copia de seguridad Almacenamiento de copias de seguridad con redundancia local
 Almacenamiento de copias de seguridad con redundancia de zona

Fuente: Elaboración propia.

4) Base de datos Rating_Device_Base_02:

Llamada "Rating_Device_Base_02" utiliza la convención camel y su nombre es fácilmente reconocible, lo que promueve la consistencia y estandarización en la denominación de los recursos, facilitando su gestión y reduciendo errores. Esta práctica se enmarca en la taxonomía de gobierno de datos (ver Ilustración 11). La

base de datos contiene información detallada del servicio, llaves de dispositivo para identificar el movimiento del usuario en la base de EventChannel y estados de servicio, incluyendo clientes activos y suspendidos, lo que ayuda a definir el total de servicios disponibles para la medición del rating.

Ilustración 11. Creación de base de datos Rating_Devices_Base_02 en nube con marcado indicado en la política Nro. 2.

Inicio > psseprycan >

Crear base de datos SQL

Microsoft

Básico Redes Seguridad Configuración adicional Etiquetas Revisar y crear

Cree una base de datos SQL con la configuración que prefiera. Complete la pestaña de configuración básica y, a continuación, vaya a Revisar y crear para efectuar el aprovisionamiento con valores predeterminados automáticos, o bien visite cada pestaña para personalizarlos. [Más información](#)

¡Sabía que los nuevos usuarios de Azure pueden crear una Azure SQL Database gratuita y usarla durante 12 meses con una cuenta gratuita de Azure? [Más información](#)

Detalles del proyecto

Seleccione la suscripción para administrar recursos implementados y los costes. Use los grupos de recursos como carpetas para organizar y administrar todos los recursos.

Suscripción

Grupo de recursos

Detalles de la base de datos

Indique la configuración necesaria para esta base de datos, incluida la selección de un servidor lógico y la configuración de los recursos de proceso y almacenamiento.

Nombre de la base de datos *

Servidor

¿Quiere usar un grupo elástico de SQL? Sí No

Proceso y almacenamiento * **Básico**
Almacenamiento: 2 GB
[Configurar base de datos](#)

Redundancia del almacenamiento de copias de seguridad

Elija el modo de replicación de las copias de seguridad de PITR y LTR. La restauración geográfica o la posibilidad de recuperación tras una interrupción regional solo están disponibles si se ha seleccionado el almacenamiento con redundancia geográfica.

Fuente: Elaboración propia.

5) Base de datos Rating_Events_Base_03

Llamada "Rating_Events_Base_03" para el proyecto sigue la convención camel en su nombre, lo que facilita su identificación y administración. Esta base de datos contiene información sobre eventos desencadenados por el televidente en los dispositivos instalados por Mayha, como el decodificador y el control remoto.

Debido a la gran cantidad de eventos registrados minuto a minuto, los datos almacenados en esta base de datos crecen en profundidad, alcanzando millones de registros (Ver Ilustración 12).

Ilustración 12. Creación de base de datos Rating_Events_Base_03 en nube con marcado indicado en la política Nro. 2.

Crear base de datos SQL ...
Microsoft

¡Sabía! que los nuevos usuarios de Azure pueden crear una Azure SQL Database gratuita y usarla durante 12 meses con una cuenta gratuita de Azure? [Más información](#)

Detalles del proyecto

Seleccione la suscripción para administrar recursos implementados y los costes. Use los grupos de recursos como carpetas para organizar y administrar todos los recursos.

Suscripción

Grupo de recursos

Detalles de la base de datos

Indique la configuración necesaria para esta base de datos, incluida la selección de un servidor lógico y la configuración de los recursos de proceso y almacenamiento.

Nombre de la base de datos *

Servidor

¿Quiere usar un grupo elástico de SQL? Sí No

Proceso y almacenamiento *
Almacenamiento: 2 GB
[Configurar base de datos](#)

Redundancia del almacenamiento de copias de seguridad

Elija el modo de replicación de las copias de seguridad de PITR y LTR. La restauración geográfica o la posibilidad de recuperación tras una interrupción regional solo están disponibles si se ha seleccionado el almacenamiento con redundancia geográfica.

Redundancia de almacenamiento de copia de seguridad Almacenamiento de copias de seguridad con redundancia local
 Almacenamiento de copias de seguridad con redundancia de zona
 Almacenamiento de copias de seguridad con redundancia geográfica

Fuente: Elaboración propia.

Se crean todos los recursos y objetos de base de datos anteriores con la marcación solicitada por Mayha en la política de Gobierno de Datos (Ver Anexo 5). También se ejecutan los cargues de la información entregada por Mayha en el servidor de Azure (Ver Ilustración 13).

Ilustración 13. Muestra del Ambiente con las BD creados.

The screenshot shows the Azure SQL Server Management Studio interface. The left sidebar contains navigation options like 'Información general', 'Registro de actividad', 'Control de acceso (IAM)', 'Etiquetas', 'Diagnóstico y solución de problemas', 'Inicio rápido', 'Configuración', 'Azure Active Directory', 'SQL Database', 'Grupos elásticos de SQL', 'Cuota de DTU', 'Propiedades', 'Bloqueos', 'Administración de datos', 'Copias de seguridad', 'Bases de datos eliminadas', 'Grupos de conmutación por error', 'Historial de importación y exportación', and 'Seguridad'. The main content area shows the configuration page for a SQL Server instance. It includes a search bar, a list of actions (Crear base de datos, Nuevo grupo elástico, etc.), and a list of characteristics (Administrador de Active Directory, Microsoft Defender para SQL, Ajuste automático, Auditoría, Grupos de conmutación por error, Cifrado de datos transparente). Below this, a table lists the available databases:

Nombre	Tipo	Estado	Plan de tarifa
Rating_Devices_Base_02	Base de datos SQL	En línea	Básico
Rating_EPG_Base_01	Base de datos SQL	En línea	Básico
Rating_Events_Base_03	Base de datos SQL	En línea	Básico

Fuente: Elaboración propia.

Azure es una plataforma en la nube que ofrece una variedad de herramientas para la gestión de datos, incluyendo la importación de archivos bacpac en Azure permite la migración de bases de datos de un entorno local a la nube de Azure con facilidad y rapidez. En el caso de la base de datos **Rating_Devices_Base_02**, se creó en la nube de Azure y Mayha entregó el archivo **Device.bacpac**, el cual contiene la estructura de la base de datos del dispositivo que se utilizará para recopilar datos de rating. Para importar este archivo en la base de datos **Rating_Devices_Base_02**, se pueden utilizar herramientas de Azure como SQL Server Management Studio o el Portal de Azure, que permiten la importación de archivos bacpac con unos pocos clics (Ver Ilustración 14).

En cuanto al archivo **EPG.bacpac**, se trata de la estructura de la base de datos que contiene la información relacionada con la programación de la guía electrónica de programación (EPG). Para importar este archivo en la base de datos **Rating_EPG_Base_01**, se puede seguir el mismo proceso descrito anteriormente (Ver Ilustración 15).

Por último, el archivo **Events.bacpac** contiene la información recopilada por Mayha a través de los dispositivos instalados por ellos, como decodificadores y controles remotos. Para importar este archivo en la base de datos **Rating_Events_Base_03**, se puede utilizar el mismo proceso de importación de archivos bacpac en herramientas de Azure como SQL Server Management Studio o el Portal de Azure (Ver Ilustración 16).

De esta manera, se pueden integrar los datos de eventos en la base de datos correspondiente para su posterior procesamiento y análisis (Ver Ilustración 17).

Ilustración 14. Importe del Archivo Device.bacpac a Rating_Device_Base_02.

Data-tier Application Wizard

Step 2: Select Import Bacpac Settings

1

2

3

File Location *

Target Server *

Target Database *

Fuente: Elaboración propia.

Ilustración 15. Importe del Archivo EPG.bacpac a Rating_EPG_Base_01.

Data-tier Application Wizard

Step 2: Select Import Bacpac Settings

1

2

3

File Location *

Target Server *

Target Database *

Fuente: Elaboración propia.

Ilustración 16. Importe del Archivo Events.bacpac a Rating_Evens_Base_03.

Data-tier Application Wizard

Step 2: Select Import Bacpac Settings

1

2

3

File Location *

Target Server *

Target Database *

Fuente: Elaboración propia.

Ilustración 17. Resultado Exitoso de la importación.

PROBLEMS	OUTPUT	TERMINAL	TASKS
✓	Import bacpac succeeded	psserprycan.database.windows.net,1433 Rating_Events_Base_03	4:17:05 p. m. - 4:20:47 p. m. (00:03:41)
✓	Import bacpac succeeded	psserprycan.database.windows.net,1433 Rating_EPG_Base_01	4:16:08 p. m. - 4:19:28 p. m. (00:03:20)
✓	Import bacpac succeeded	psserprycan.database.windows.net,1433 Rating_Devices_Base_02	4:14:11 p. m. - 4:16:27 p. m. (00:02:16)

Fuente: Elaboración propia.

Como resultado del cargue y la consulta de datos, se generó un Modelo de datos Lógico (Ver Ilustración 18) y un Modelo de datos Físico (Ver Ilustración 19), que permitió identificar los elementos clave que se utilizan en el cálculo de los indicadores para el Dashboard.

Los modelos cuentan con 8 entidades que definen claramente los principales componentes de la información: Canales, Categorías, Genero, Clasificación, Idioma, Tipo Show, Programas y todo el histórico de la guía electrónica de programación que contiene los programas de tv que efectivamente se transmitieron a través de los canales y horarios definidos.

También, se pueden evidenciar las relaciones entre estas entidades en el modelo físico descrito como llaves foráneas que es un elemento dentro de la tabla que permite identificar de manera única una entidad de un conjunto de entidades.

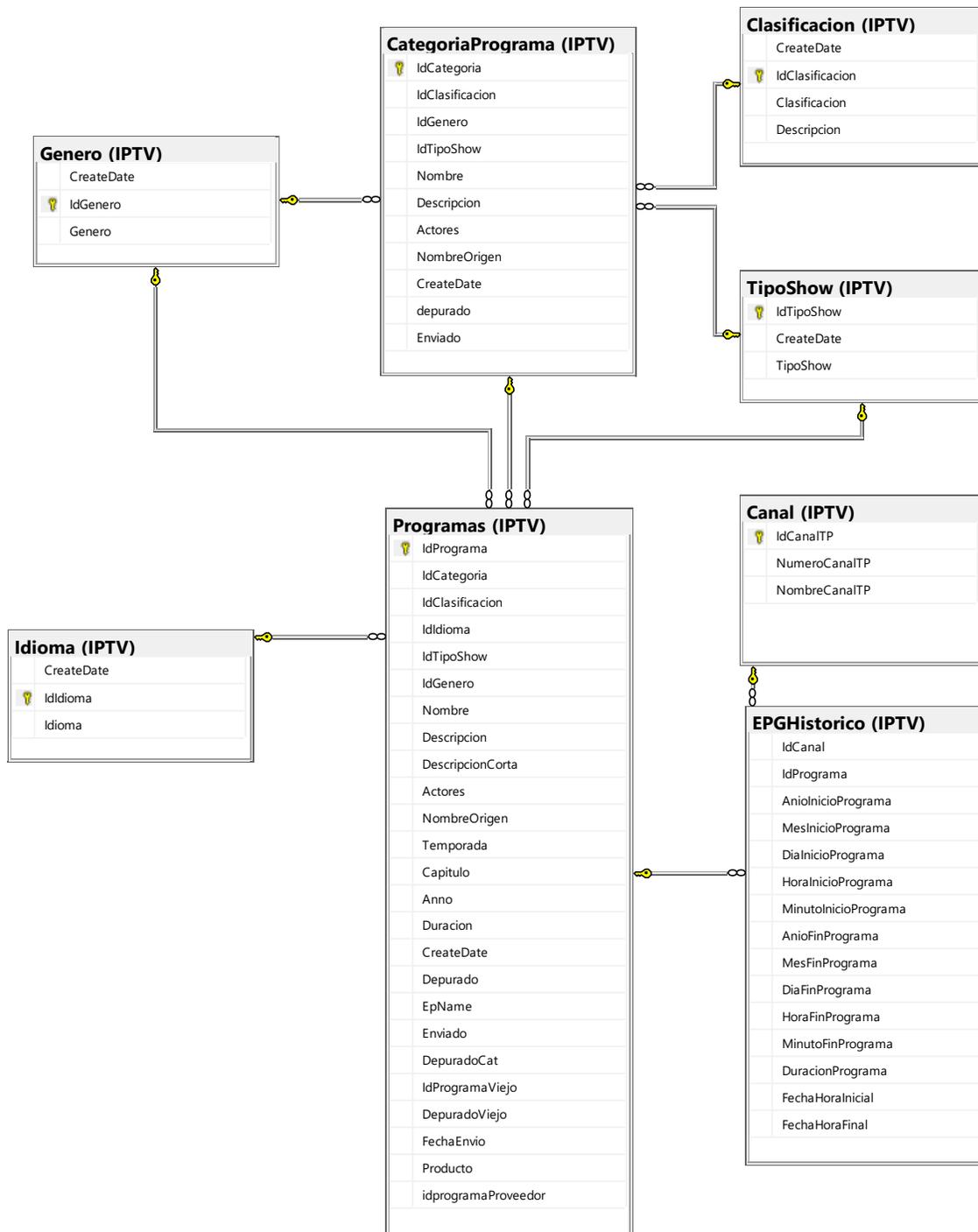
Se puede evidenciar también como estas llaves foraneas del presente modelo facilita la conexión entre las tablas en la base de datos.

Tomando en cuenta que las claves primarias hacen referencia a una identificación única para un registro, con la clave foránea no siempre se da este aspecto, es decir, que no siempre es única.

Las llaves foráneas de estas tablas necesariamente son la llave primaria de la tabla de dónde vienen y facilitan el relacionamiento en las consultas al motor de bases de datos de las tablas implementadas en Azure.

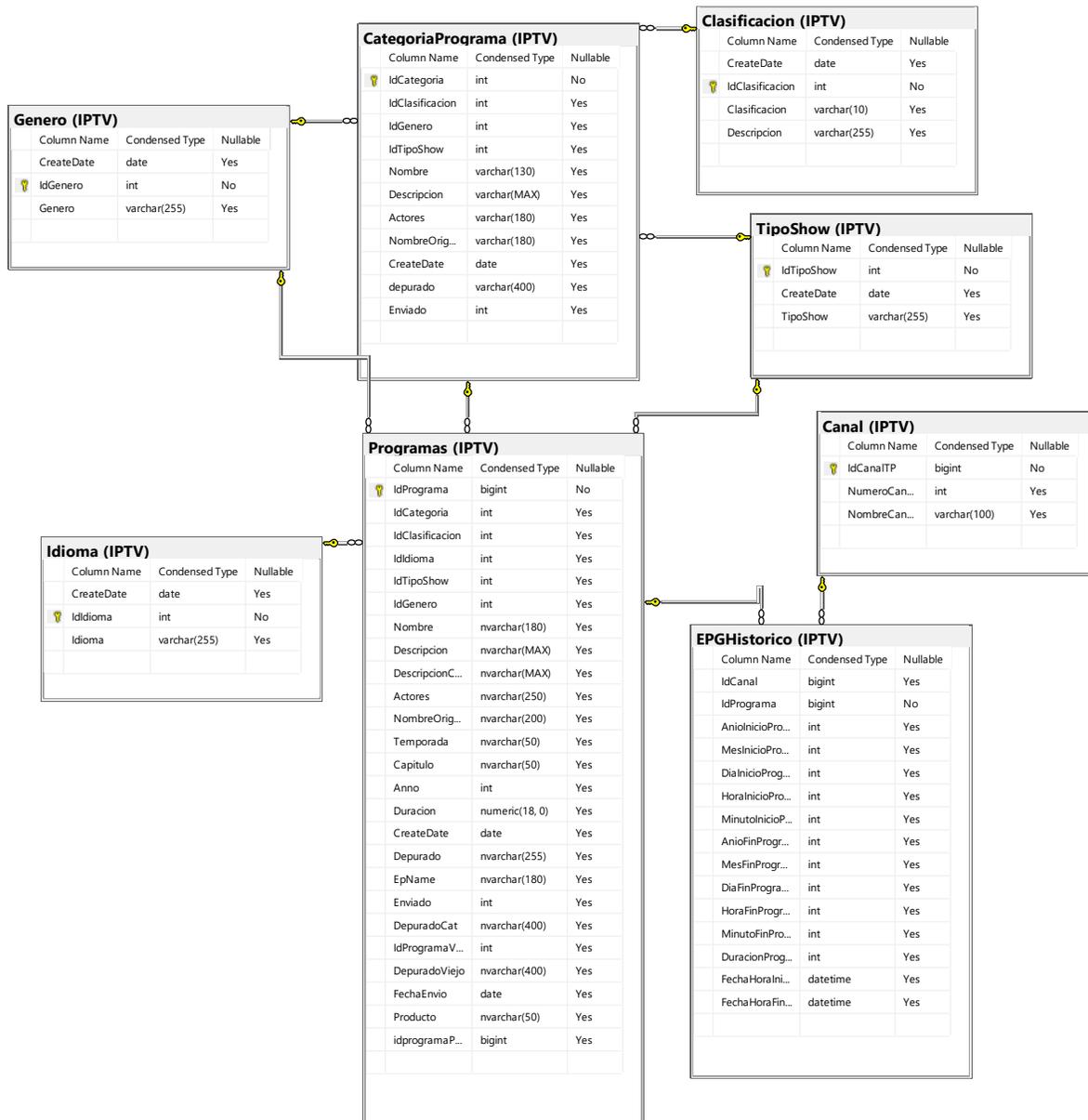
Por último, esta información junto con los datos de la base de eventos y devices, se procesa en un nuevo modelo para el análisis y consultas desde el dashboard de audiencias.

Ilustración 18. Modelo lógico de base de datos EPG



Fuente: Elaboración propia.

Ilustración 19. Modelo Físico.



Fuente: Elaboración Propia

A su vez, dentro de la identificación de los datos, contamos con dos tablas: "Events" que pertenece a la base de datos "Rating_Events_Base_03" (Ver Ilustración 20) y "InventoryDevice" (Ver Ilustración 21) que pertenece a la base de datos "Rating_Device_Base_02". La descripción detallada de estas bases de datos se encuentra en el punto 5.4.

Ilustración 20. Tabla Events

Event (EVENTS)		Event (EVENTS)		
Column Name	Condensed Type	Nullable		
IdEvent	int	No		
TypeDevice	varchar(20)	No		
MACDevice	varchar(17)	No		
SerialDevice	int	No		
StartNameDay	varchar(10)	No		
StartTuningTime	datetime	No		
EndTuningTime	datetime	No		
EndNameDay	varchar(10)	No		
Duration	int	No		
TypeTuning	varchar(10)	No		
StripStart	varchar(35)	No		
StripStartReference	int	Yes		
StripEndReference	int	Yes		
StripEnd	varchar(35)	No		
StripEndStartReference	int	Yes		
StripEndEndReference	int	Yes		
Week	int	No		

Fuente: Elaboración Propia

Ilustración 21. Tabla InventoryDevice.

InventoryDevice (dbo)		InventoryDevice (dbo)		
Column Name	Condensed Type	Nullable		
MACDevice	varchar(20)	Yes		
SerialDevice	varchar(20)	No		
TypeDevice	varchar(25)	Yes		
CodCity	varchar(10)	Yes		
Neighborhood	varchar(60)	Yes		
EconomicCode	int	Yes		

Fuente: Elaboración Propia

5.6 Alistamiento de los datos:

Una vez se inicia la extracción detallada en la definición anterior; a través de consultas se comienza a verificar las reglas que debe cumplir la información, se da inicio al proceso de transformación de la data aplicando limpieza y preparación que quedaran almacenada en la política de gobierno de datos para este proyecto (Ver Anexo 6).

El proceso de normalización es una práctica común en la gestión de bases de datos, permitiendo afinar la estructura de la base de datos y asegurar la integridad de los datos almacenados. En el caso específico de la BD Rating_EPG_Base_01, se identificó la necesidad de normalizar la tabla EPGHistorico para mejorar su eficiencia y funcionalidad.

Para ello, se realizó una revisión detallada de la tabla y se identificaron los problemas existentes, como la falta de un tipo de dato adecuado y la ausencia de llave primaria y relaciones con otras tablas (Ver Ilustración 22). Con base en esta información, se implementaron las correcciones necesarias para normalizar la tabla, incluyendo la adición de una llave primaria y la creación de restricciones de integridad referencial para establecer relaciones con otras tablas de la base de datos (Ver Ilustración 23).

Este proceso de normalización permitió mejorar la calidad de los datos almacenados en la tabla EPGHistorico y asegurar su consistencia y coherencia con el resto de la base de datos. Asimismo, contribuyó a optimizar la eficiencia y rendimiento de la BD Rating_EPG_Base_01 en su conjunto.

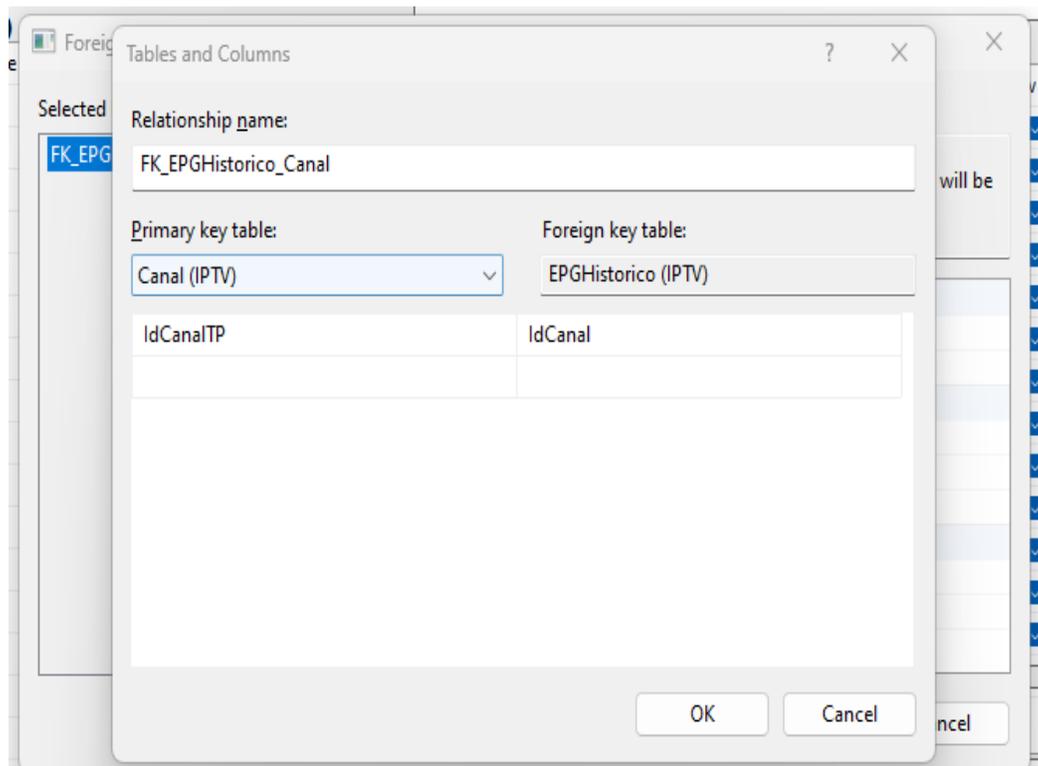
Ilustración 22. Tabla EPGHistorico sin relación.

Column Name	Data Type	Allow Nulls
IdCanal	int	<input checked="" type="checkbox"/>
IdPrograma	bigint	<input checked="" type="checkbox"/>
AnioInicioPrograma	int	<input checked="" type="checkbox"/>
MesInicioPrograma	int	<input checked="" type="checkbox"/>
DiaInicioPrograma	int	<input checked="" type="checkbox"/>
HoraInicioPrograma	int	<input checked="" type="checkbox"/>
MinutoInicioPrograma	int	<input checked="" type="checkbox"/>
AnioFinPrograma	int	<input checked="" type="checkbox"/>
MesFinPrograma	int	<input checked="" type="checkbox"/>
DiaFinPrograma	int	<input checked="" type="checkbox"/>
HoraFinPrograma	int	<input checked="" type="checkbox"/>
MinutoFinPrograma	int	<input checked="" type="checkbox"/>
DuracionPrograma	int	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

Column Name	Data Type	Allow Nulls
IdCanalTP	bigint	<input type="checkbox"/>
NumeroCanalTP	int	<input checked="" type="checkbox"/>
NombreCanalTP	varchar(100)	<input checked="" type="checkbox"/>
		<input type="checkbox"/>

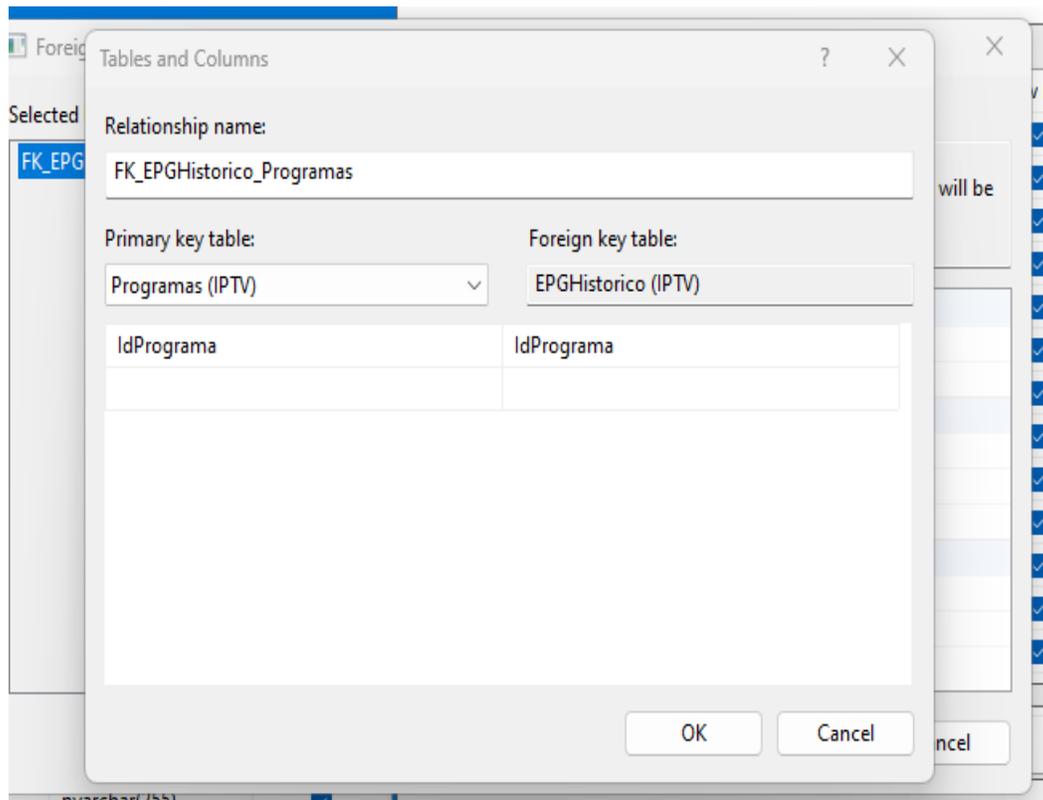
Fuente: Elaboración Propia

Ilustración 23. Creación de restricciones



Fuente: Elaboración Propia

Además de la normalización de la BD Rating_EPG_Base_01 mencionada anteriormente, se identificó que la tabla necesitaba tener una relación con la tabla EPGHistorico para mejorar la integridad referencial de la base de datos. Para lograr esto, se implementó una restricción de entidad referencial para asegurar que cualquier registro en la tabla de programas tenga un correspondiente en la tabla de EPGHistorico. Esto permite mantener la consistencia y precisión de los datos en la base de datos, asegurando que la información esté disponible y sea confiable para su uso en el proyecto (Ver Ilustración 24).

Ilustración 24. Creación de restricciones programas.

Fuente: Elaboración Propia

Es crucial realizar un proceso de verificación y limpieza de los datos en la base de datos Rating_EPG_Base_01 para cumplir con la regla establecida de no tomar en cuenta para las mediciones del rating aquellos programas que cumplan con ciertos criterios. Esto implica filtrar y eliminar los programas que contengan los nombres "Infomercial" o "Infomerciales" en el campo IPTV.Programas.Nombre durante los meses de octubre, noviembre y diciembre del año 2022. Y enero, febrero y marzo del año 2023. Este proceso se lleva a cabo mediante un script que automatiza las tareas repetitivas y aplica reglas de negocio específicas para garantizar la coherencia y precisión de los datos. La verificación y limpieza de los datos son esenciales para asegurar la calidad e integridad de la información, lo que a su vez permite tomar decisiones basadas en los datos obtenidos (Ver Ilustración 25).

Ilustración 25. Consulta de verificación de existencia de programas de infomerciales.

```

USE EPG
-----este conjunto de datos no debe pasar a medicion de rating -----
SELECT      IPTV.Canal.NumeroCanalTP, IPTV.Canal.NombreCanalTP, IPTV.EPGHistorico.DuracionPrograma,
            IPTV.Programas.Nombre, IPTV.EPGHistorico.AnioInicioPrograma, IPTV.EPGHistorico.MesInicioPrograma,
            IPTV.EPGHistorico.DiaInicioPrograma, IPTV.EPGHistorico.HoraInicioPrograma
FROM        IPTV.EPGHistorico INNER JOIN
            IPTV.Programas ON IPTV.EPGHistorico.IdPrograma = IPTV.Programas.IdPrograma INNER JOIN
            IPTV.Canal ON IPTV.EPGHistorico.IdCanal = IPTV.Canal.IdCanalTP
WHERE       (IPTV.Programas.Nombre LIKE '%Infomercial%') and AnioInicioPrograma in ('2022', '2023') and MesInicioPrograma in (10,11,12,1,2,3)
  
```

	NumeroCanalTP	NombreCanalTP	DuracionPrograma	Nombre	AnioInicioPrograma	MesInicioPrograma	DiaInicioPrograma	HoraInicioPrograma
1	606	Nat Geo Colombia	30	Infomercial	2022	10	1	7
2	606	Nat Geo Colombia	30	Infomercial	2022	10	1	9
3	606	Nat Geo Colombia	30	Infomercial	2022	10	2	8

Fuente: Elaboración propia.

El script es una consulta SQL que busca obtener información específica de la base de datos "Rating_EPG_Base_01". En la consulta se utilizan varias tablas que se relacionan entre sí por medio de claves primarias y foráneas.

La consulta utiliza la función ROW_NUMBER() para asignar un número de fila a cada registro obtenido, el cual se ordena por el número de canal TP. Luego, se seleccionan varias columnas de las tablas "EPGHistorico", "Programas" y "Canal", como el número de canal, el nombre del canal, la duración del programa, el nombre del programa, la descripción corta, los actores y el producto.

La consulta utiliza la cláusula WHERE para filtrar los resultados. Se excluyen los programas cuyo nombre contiene la cadena "Infomercial", y se incluyen utilizando la cláusula IN de SQL solo los programas que iniciaron en los meses de octubre de 2022, noviembre de 2022, diciembre de 2022, enero de 2023, febrero de 2023 y marzo de 2023.

En resumen, el script es una consulta que busca obtener información específica de la base de datos, y utiliza filtros para obtener solo los resultados que cumplen con ciertos criterios establecidos. Una vez identificados, este mismo script servirá para descartarlos en el cargue hacia la base de datos: Rating_Staging_Base_04.

De la misma manera se realiza la comprobación e identificación de los programas con duraciones menores o iguales a 5 minutos ejecutando la consulta y constatando que efectivamente se realizaron emisiones de programas muy cortos que no deben tenerse en cuenta dentro del cálculo del rating como es descrito en la política no. 3 (Ver ilustración 26).

Ilustración 26. Verificación de la existencia de programas emitidos con duración inferior a 5 minutos.

```

-----este conjunto de datos no debe pasar-----
SELECT      IPTV.Canal.NumeroCanalTP, IPTV.Canal.NombreCanalTP, IPTV.Programas.Nombre, IPTV.EPGHistorico.DuracionPrograma
FROM        IPTV.EPGHistorico INNER JOIN
            IPTV.Canal ON IPTV.EPGHistorico.IdCanal = IPTV.Canal.IdCanalTP INNER JOIN
            IPTV.Programas ON IPTV.EPGHistorico.IdPrograma = IPTV.Programas.IdPrograma
WHERE      DuracionPrograma <= 5

-----DURATION SURFING-----

```

NumeroCanalTP	NombreCanalTP	Nombre	DuracionPrograma
758	HBO Plus Colombia	El espacio entre nosotros	5
758	HBO Plus Colombia	El espacio entre nosotros	5
722	Multicinema Colombia	Elysium	5
722	Multicinema Colombia	Pacto Final	5

Fuente: elaboración propia

Una vez identificados, este mismo script sirve para descartarlos en el cargue hacia la base de datos: **Rating_Staging_Base_04**.

5.7 Almacenamiento:

Esta etapa a medida que se va generando y aplicando los procesos de transformación la data limpia y preparada se carga en una base de datos que generalmente se denomina Staging, en este punto es posible que se sigan aplicando procesos de limpieza hasta que la información quede refinada y preparada.

Como se mencionó anteriormente, se siguen aplicando procesos de limpieza y corrección como la siguiente. Se realiza un proceso que consiste en solucionar la fragmentación que existe entre los campos AnioInicioPrograma, MesInicioPrograma, DialInicioPrograma, HoralInicioPrograma, MinutoInicioPrograma y AnioFinPrograma, MesFinPrograma, DiaFinPrograma, HoraFinPrograma, MinutoFinPrograma. Esta fragmentación puede dificultar la realización de relaciones de validación y almacenamiento. Para solucionar este problema, se realiza una concatenación de campos para crear dos nuevos campos: FechaHoralInicial y FechaHoraFin. De esta manera, se pueden manejar los rangos de fecha de manera más fácil y efectiva. Con esta corrección, se asegura que la información sea más coherente y precisa, lo que contribuye a la toma de decisiones informadas y precisas en el futuro (Ver Ilustración 27).

Ilustración 27. Desfragmentación de la fecha hora inicial y final de los programa y actualización en la base extraída.

```

SELECT IdCanal, IdPrograma, AnioInicioPrograma,
       MesInicioPrograma, DiaInicioPrograma, HoraInicioPrograma,
       MinutoInicioPrograma, AnioFinPrograma, MesFinPrograma,
       DiaFinPrograma, HoraFinPrograma, MinutoFinPrograma, DuracionPrograma
       , CONCAT(AnioInicioPrograma, '-', MesInicioPrograma, '-', DiaInicioPrograma, ':', HoraInicioPrograma, ':', MinutoInicioPrograma, ':00') FechaInicial
       , CONCAT(AnioFinPrograma, '-', MesFinPrograma, '-', DiaFinPrograma, ':', HoraFinPrograma, ':', MinutoFinPrograma, ':00') FechaFinal
FROM EPG.[IPTV].[EPGHistorico]

```

IdCanal	IdPrograma	AnioInicioPrograma	MesInicioPrograma	DiaInicioPrograma	HoraInicioPrograma	MinutoInicioPrograma	AnioFinPrograma	MesFinPrograma	DiaFinPrograma	HoraFinPrograma
1805	509853	2022	11	8	21	0	2022	11	8	21
1805	510279	2023	1	10	14	0	2023	1	10	17
1805	517909	2022	10	30	16	30	2022	10	30	17

Fuente: elaboración propia

El proceso llevado a cabo se encuadra dentro de la gestión de calidad de datos, en cuanto a la consistencia y precisión de los mismos. Se buscó solucionar una fragmentación en la estructura de los datos, con el objetivo de lograr una mayor coherencia y precisión en la información. De esta manera, se contribuyó a una gestión más efectiva y eficiente del gobierno de datos, que se enfoca en establecer políticas, estándares y procedimientos para garantizar la integridad, consistencia, precisión y accesibilidad de los datos en toda la organización.

Una vez que se han llevado a cabo los procesos de limpieza y corrección de los datos, se da paso a la creación de los ETL's. Este proceso es fundamental para la transformación y almacenamiento de los datos en la base de datos. Para ello, se utiliza la herramienta Azure Data Factory, la cual permite ejecutar consultas y transformaciones de datos de forma sencilla y eficiente.

Con la herramienta Azure Data Factory, se pueden realizar múltiples tareas, como extracción de datos de diferentes orígenes de datos, transformación de datos y carga de datos en la base de datos. Esta herramienta facilita el proceso de ETL, ya que permite automatizar los procesos de integración de datos y garantiza la integridad y calidad de estos.

Una vez que se ejecuta el proceso de ETL con la herramienta Azure Data Factory, se logra almacenar los datos de forma eficiente y estructurada en la base de datos, lo que permite su fácil acceso y análisis posterior. De esta manera, se garantiza que la información almacenada sea confiable y se pueda utilizar para la toma de decisiones.

Como primera medida se crea una consulta que realiza una búsqueda en la tabla "Rating_EPG_Base_01.IPTV.EPGHistorico" para obtener el Id del canal, el Id del programa y el Id del evento de sintonía para cada programa. La subconsulta dentro de la consulta principal busca el Id del evento de sintonía correspondiente al programa actual a partir de la tabla "EVENTS.Event" y utilizando el rango de fechas/horas del programa (FechaHoraInicial y FechaHoraFinal) como criterio de búsqueda. El resultado final es una tabla que contiene tres columnas: IdCanal, IdPrograma e IdSintonia (Ver Ilustración 28).

Ilustración 28. Consulta que relaciona los eventos vs los programas.

```

SELECT EP.IdCanal, EP.IdPrograma,
(SELECT TOP 1 IdEvent FROM [EVENTS].[Event] EC
WHERE EC.StartTuningTime Between EP.FechaHoraInicial AND EP.FechaHoraFinal)IdSintonia
FROM Rating_EPG_Base_01.[IPTV].[EPGHistorico] EP WITH (NOLOCK)

```

Fuente: elaboración propia

Dado el volumen considerable de datos obtenidos a través de la consulta anterior, se hace necesario la creación de un ETL llamado CP_Up_EventTuning en la herramienta Azure Data Factory. Este proceso tomará los datos resultantes de la consulta y, mediante la actividad "Copiar Datos", insertará los mismos en la tabla "EventTuning". Como resultado, se obtiene la relación existente entre la actividad del usuario y la programación emitida por los canales en el periodo de tiempo definido durante la etapa de alistamiento de datos. De esta forma, se podrá obtener información relevante y valiosa para la toma de decisiones (Ver Ilustración 29-33).

Ilustración 29. Creación de Etl y actividad Copiar Datos.

The screenshot displays the Azure Data Factory 'Copiar datos' activity configuration. The activity name is 'CP_Up_EventTuning'. The description is 'Pipeline que actualiza la tabla, [EVENTS].[EventTuning] de la Bd Rating_Events_Base_03'. The source data set is 'Data_Origen_EPG_01'. The query is:


```

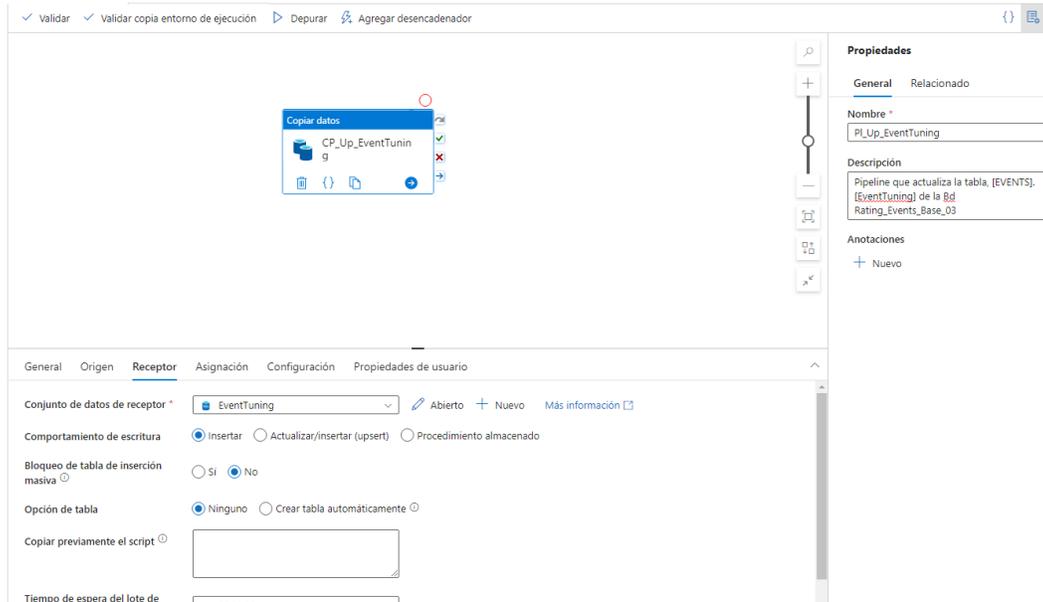
SELECT EP.IdCanal, EP.IdPrograma,
(SELECT TOP 1 IdEvent FROM [EVENTS].[Event] EC
WHERE EC.StartTuningTime Between EP.FechaHoraInicial AND EP.FechaHoraFinal)IdSintonia
FROM Rating_EPG_Base_01.[IPTV].[EPGHistorico] EP
WITH (NOLOCK)

```

 The query timeout is set to 120 minutes. The interface also shows the 'Propiedades' (Properties) pane on the right and the 'Actividades' (Activities) pane on the left.

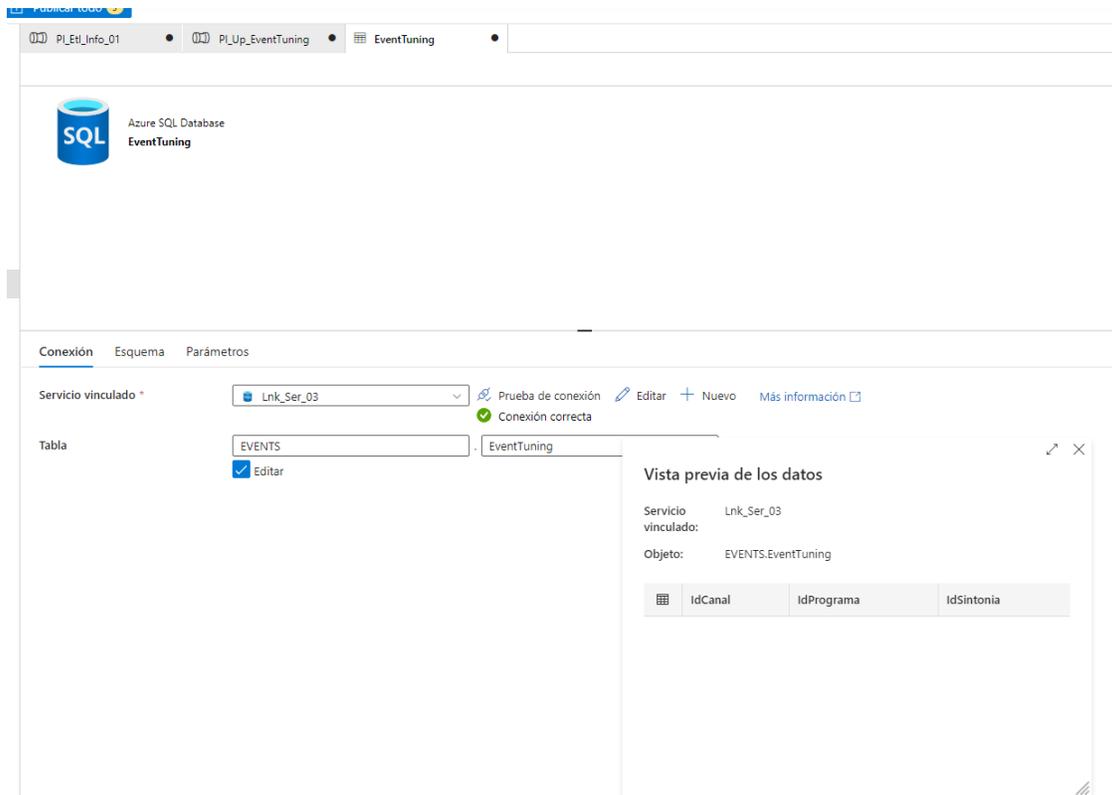
Fuente: elaboración propia

Ilustración 30. Creación de objeto destino.



Fuente: elaboración propia

Ilustración 31. Configuración de Tabla destino EventTuning.



EventTuning (EVENTS)	
	IdCanal
	IdPrograma
	IdSintonia

Fuente: elaboración propia

Ilustración 32. Evidencia de procesamiento de ETL.

Detalles actualizar

Sugerencias de optimización de rendimiento:
Fuente habilitada para partición; para lograr el mejor rendimiento al copiar grandes cantidades de datos de esta fuente, se sugiere utilizar el paralelo integrado configurando las opciones de partición en la configuración de la fuente. Consulte este documento.

Puede obtener más información sobre los detalles de rendimiento de la copia aquí.

Identificación de ejecución de actividad: 13159058-68b4-4bb3-9e7f-c9d1e1396036



Base de datos Azure SQL
Región: Este de EE.UU.

en curso

→



Base de datos Azure SQL
Región: Este de EE.UU.

Región de Azure IR: Este de EE.UU.

Datos leídos:	232.52KB	Datos escritos:	0 bytes
Filas leídas:	11.626	Filas escritas:	0
Conexiones máximas:	1	Conexiones máximas:	2

Duración de la copia: 10:16:18
Rendimiento: 10 bytes/segundo

Base de datos Azure SQL → Base de datos Azure SQL

hora de inicio: 2023-05-01T03:29:53.1890962Z
DIU utilizados: 4
Copias en paralelo usadas: 1
Duración: 10:16:18

Detalles	Duración del trabajo	Duración total
Reajuste salarial		00:00:03
Transferir	Tiempo hasta el primer byte: 00:36:52 Leyendo desde el origen: 01:05:16 Escribiendo en el receptor: 00:00:00	10:16:13

Comprobación de coherencia de datos: sin verificar

Fuente: elaboración propia

Ilustración 33. Resultado de la ETL CP_Up_EventTuning.

	IdCanal	IdPrograma	IdSintonia
1	1805	429732	NULL
2	1805	429732	NULL
3	1805	429732	19763
4	1805	429732	317305
5	1805	429732	317305
6	1805	429732	380502
7	1805	429732	NULL
8	1805	429732	NULL
9	1805	429732	NULL
10	1805	429732	NULL
11	1805	429732	372793
12	1805	429732	399302
13	1805	429732	463870
14	1805	429732	NULL
15	1805	429732	409723
16	1805	429732	NULL
17	1805	429732	473447
18	1805	429732	NULL
19	1805	429732	473105

Fuente: elaboración propia

5.8 Procesamiento:

Esta etapa se refiere al análisis de los datos para obtener resultados útiles. Esto incluye la creación de nuevas estructuras que se diseñan para ayudar a la analítica de los datos como modelos estrella, copo de nieve, cubos de datos e incluye la codificación de medidas particulares que ayuden a generar los cálculos de los indicadores de los programas de televisión.

Para ello se realiza la creación de la base de datos (bd) `Rating_Staging_Base_04`, donde se almacenará la data limpia y procesada, este punto se realiza sobre la nube de Azure (Ver Ilustración 34).

Ilustración 34. Creación de bd Rating_Staging_Base_04.

Crear base de datos SQL

Microsoft

¿Sabía que los nuevos usuarios de Azure pueden crear una Azure SQL Database gratuita y usarla durante 12 meses con una cuenta gratuita de Azure? [Más información](#)

Detalles del proyecto

Seleccione la suscripción para administrar recursos implementados y los costes. Use los grupos de recursos como carpetas para organizar y administrar todos los recursos.

Suscripción

Grupo de recursos

Detalles de la base de datos

Indique la configuración necesaria para esta base de datos, incluida la selección de un servidor lógico y la configuración de los recursos de proceso y almacenamiento.

Nombre de la base de datos *

Servidor

¿Quiere usar un grupo elástico de SQL? Sí No

Proceso y almacenamiento * **Básico**
Almacenamiento: 2 GB
[Configurar base de datos](#)

Redundancia del almacenamiento de copias de seguridad

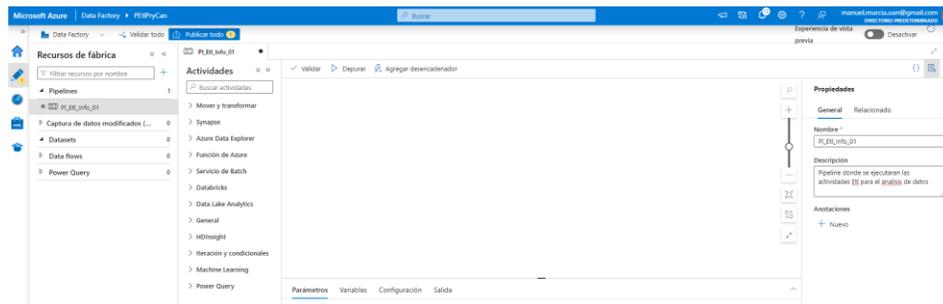
Elija el modo de replicación de las copias de seguridad de PITR y LTR. La restauración geográfica o la posibilidad de recuperación tras una interrupción regional solo están disponibles si se ha seleccionado el almacenamiento con redundancia geográfica.

Redundancia de almacenamiento de copia de seguridad Almacenamiento de copias de seguridad con redundancia local
 Almacenamiento de copias de seguridad con redundancia de zona
 Almacenamiento de copias de seguridad con redundancia geográfica

Fuente: elaboración propia

El paso siguiente procede a crear un segundo ETL llamado PL_Etl_Info_01 en la herramienta Azure Data Factory. Este ETL se encargará de procesar los datos ya limpios. En este proceso se ejecutarán diversas transformaciones de datos para lograr una mejor organización y presentación de estos. Con el fin de optimizar el uso de recursos y mejorar la eficiencia en el procesamiento de datos, se implementaron técnicas de paralelismo en la ejecución de las tareas. De esta manera, se logra obtener información de mayor calidad y precisión, lo que contribuirá a una toma de decisiones más acertada (Ver Ilustración 35).

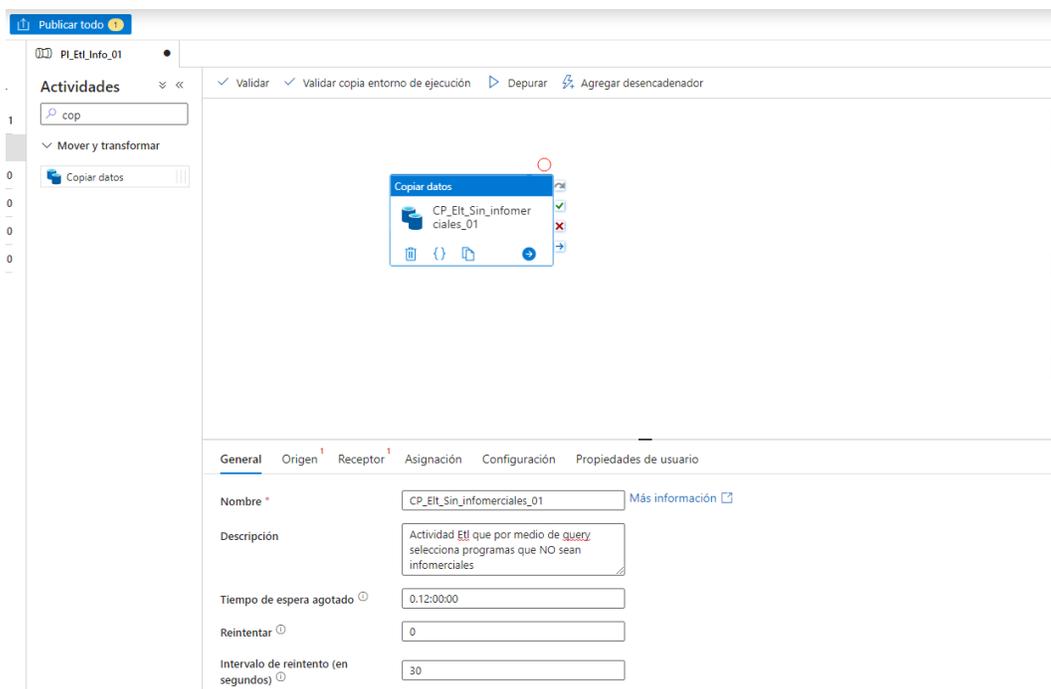
Ilustración 35. Creación de ETL PL_Etl_Info_01.



Fuente: elaboración propia

Dentro del ETL se configura la actividad "Copiar Datos" la cual llamaremos CP_Etl_Sin_Infomerciales_01. En esta actividad es necesario indicar los datos fuente que se procesarán en la ejecución. Para este caso, la fuente es la base de datos Rating_EPG_Base_01, a la cual se accederá mediante un linked server. Esta base de datos es sobre la cual ya hemos realizado procesos de normalización y calidad de datos en las etapas anteriores (Ver Ilustración 36, 37).

Ilustración 36. Configuración de ETL PL_Etl_Info_01.



Fuente: elaboración propia

Ilustración 37. Creación de linked server a bd Rating_EPG_Base_01.

Nuevo servicio vinculado

Azure SQL Database [Más información](#)

Nuevo servicio vinculado

Nombre *
Lk_Ser_01

Descripción
Linkend Service que conectara con la BD Rating_EPG_Base_01

Conectar mediante Integration Runtime * ⓘ
AutoResolveIntegrationRuntime

Cadena de conexión Azure Key Vault

Método de selección de cuenta ⓘ
 Desde una suscripción de Azure Indicar manualmente

Suscripción de Azure
Azure for Students (a8cd9e31-38d0-4562-8a36-5c17a1f01b6f)

Nombre del servidor *
psserprycan

Nombre de la base de datos *
Rating_EPG_Base_01

Tipo de autenticación *
Autenticación SQL

Nombre de usuario *
Administrador

Contraseña Azure Key Vault

Contraseña *
.....

Siempre cifradas ⓘ

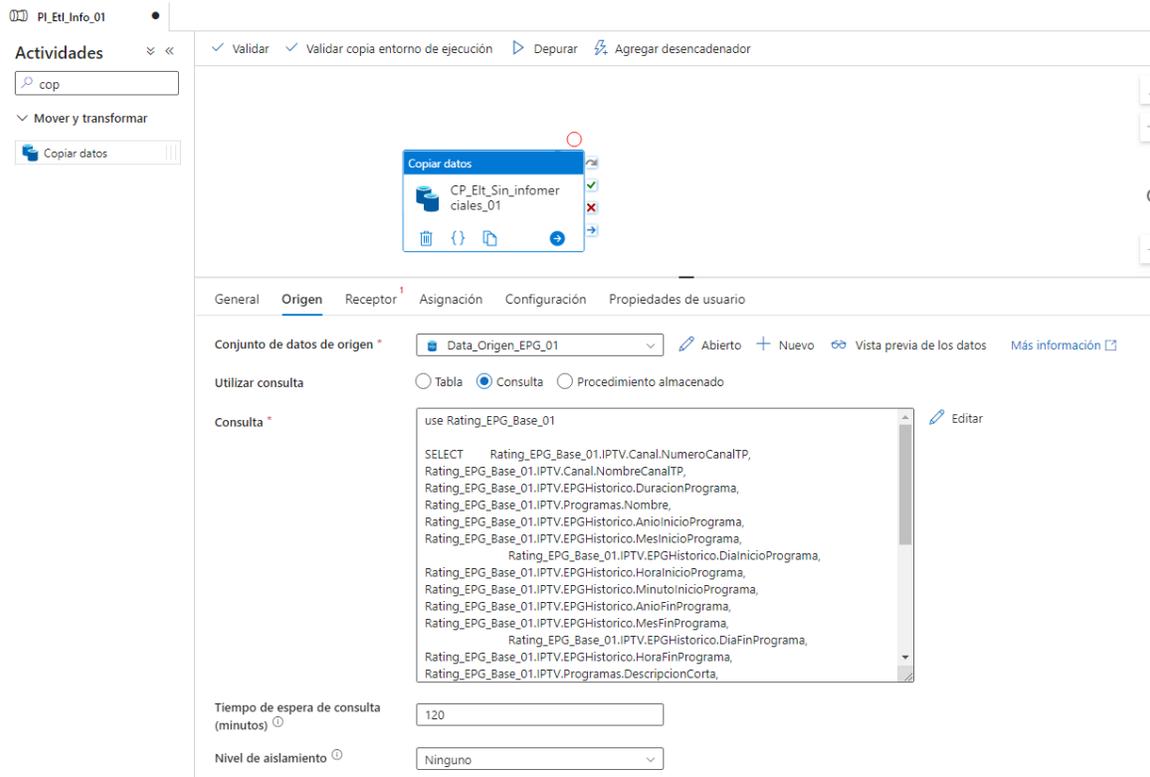
Crear Cancelar

Conexión correcta
Prueba de conexión

Fuente: elaboración propia

Es importante destacar que en la fuente de origen se han definido los scripts SQL (ver Ilustración 25 y 26) que fueron depurados previamente con base en las reglas de negocio establecidas por la empresa Mayha. Esto garantiza que los datos que se procesen en el ETL sean confiables y coherentes con los objetivos del proyecto. La depuración de los scripts es un paso fundamental en el proceso de normalización y calidad de datos, ya que ayuda a evitar posibles errores e inconsistencias en los resultados obtenidos (Ver Ilustración 38).

Ilustración 38. Configuración de la actividad CP_Etl_Sin_Infomerciales_01.



The screenshot displays the configuration for the 'Copiar datos' activity in the 'Origen' (Source) tab. The 'Conjunto de datos de origen' (Data source) is set to 'Data_Origen_EPG_01'. The 'Utilizar consulta' (Use query) option is selected, and the 'Consulta' (Query) field contains the following SQL statement:

```
use Rating_EPG_Base_01
SELECT      Rating_EPG_Base_01.IPTV.Canal.NumeroCanalITP,
           Rating_EPG_Base_01.IPTV.Canal.NombreCanalITP,
           Rating_EPG_Base_01.IPTV.EPGHistorico.DuracionPrograma,
           Rating_EPG_Base_01.IPTV.Programas.Nombre,
           Rating_EPG_Base_01.IPTV.EPGHistorico.AnioInicioPrograma,
           Rating_EPG_Base_01.IPTV.EPGHistorico.MesInicioPrograma,
           Rating_EPG_Base_01.IPTV.EPGHistorico.DialInicioPrograma,
           Rating_EPG_Base_01.IPTV.EPGHistorico.HoraInicioPrograma,
           Rating_EPG_Base_01.IPTV.EPGHistorico.MinutoInicioPrograma,
           Rating_EPG_Base_01.IPTV.EPGHistorico.AnioFinPrograma,
           Rating_EPG_Base_01.IPTV.EPGHistorico.MesFinPrograma,
           Rating_EPG_Base_01.IPTV.EPGHistorico.DiaFinPrograma,
           Rating_EPG_Base_01.IPTV.EPGHistorico.HoraFinPrograma,
           Rating_EPG_Base_01.IPTV.Programas.DescripcionCorta,
```

The 'Tiempo de espera de consulta (minutos)' (Query timeout) is set to 120, and the 'Nivel de aislamiento' (Isolation level) is set to 'Ninguno'.

Fuente: elaboración propia

Se realiza la configuración del destino para almacenar los datos procesados. En este caso, se crea un nuevo linked server sobre la base de datos Rating_Staging_Base_04 y la tabla SabanaReporte. Esto permite realizar la inserción de la información. Además, se agrega un parámetro "DROP TABLE [dbo].[SabanaReporte]" que realiza la limpieza de la tabla para recibir nuevos registros cada vez que se ejecute el ETL. Es importante mencionar que esta tabla se utiliza para generar reportes y análisis de datos relevantes para la toma de decisiones en la empresa (Ver Ilustración 39, 40).

Ilustración 39. Creación de linked server a bd Rating_Staging_Base_04.

Fuente: elaboración propia

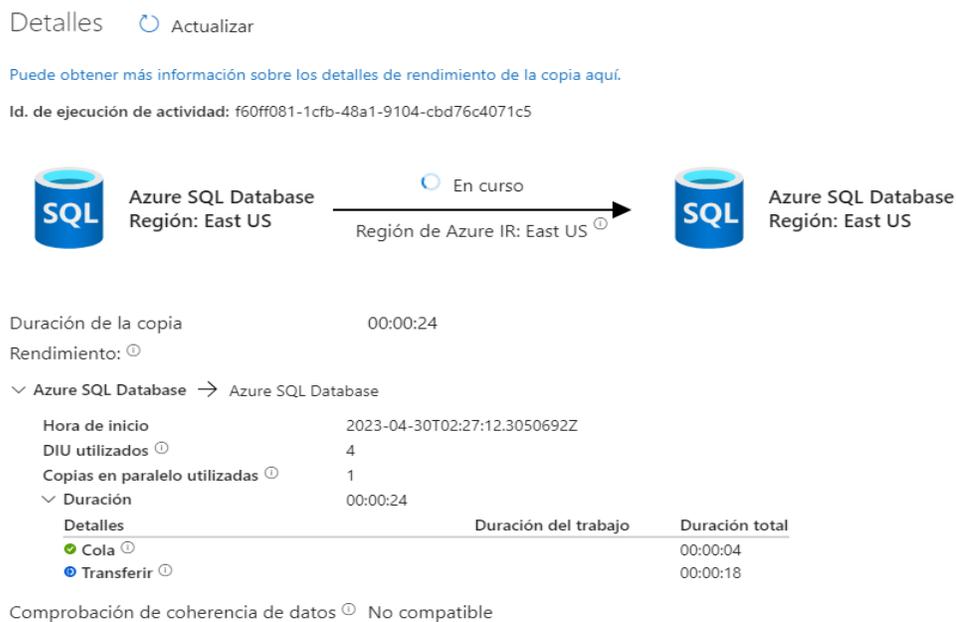
Ilustración 40. Configuración destino tabla SabanaReporte.

Fuente: elaboración propia

Durante esta etapa, también se llevan a cabo actividades de monitoreo que permiten obtener información sobre la duración del proceso, la cantidad de registros procesados y los recursos utilizados. Esta data es de gran importancia, ya que permite visualizar de manera clara y precisa si el ETL ha sido exitoso y si se han enriquecido los registros en el destino.

Estas actividades de monitoreo pueden ser realizadas por medio de herramientas de seguimiento y supervisión, como por ejemplo Azure Monitor, que permite obtener estadísticas detalladas sobre el rendimiento y la eficiencia del proceso. De esta manera, se pueden identificar posibles cuellos de botella, optimizar el proceso y mejorar la calidad de los datos procesados (Ver Ilustración 41, 42, 43).

Ilustración 41. Monitoreo sobre el ETL PL_Etl_Info_01.



Fuente: elaboración propia

Ilustración 42. Finalización exitosa del ETL Pl_Etl_Info_01.

Detalles Actualizar

Sugerencias de optimización de rendimiento:
Sink Azure SQL Database: The DTU utilization was high during the copy activity run. To achieve better performance, you are suggested to scale the database to a higher tier than the current 5 DTUs. Consulte este documento .

Puede obtener más información sobre los detalles de rendimiento de la copia aquí.

Id. de ejecución de actividad: f60f081-1c1b-48a1-9104-cbd76c4071c5

Correcto
Región de Azure IR: East US

Azure SQL Database
Región: East US

Datos leídos: 129,903 MB
Filas leídas: 731,616
Conexiones máximas: 1

Azure SQL Database
Región: East US

Datos escritos: 129,903 MB
Filas escritas: 731,616
Conexiones máximas: 3

Duración de la copia: 00:07:03
Rendimiento: 319,172 KB/s

Azure SQL Database → Azure SQL Database

Hora de inicio: 2023-04-30T02:27:12.3050692Z
DIU utilizados: 4
Copias en paralelo utilizadas: 1

Duración: 00:07:03

Detalles	Duración del trabajo	Duración total
Cola		00:00:04
Script anterior a la copia		00:00:00
Transferir	Tiempo hasta el primer byte: 00:00:11 Leyendo desde el origen: 00:00:26 Escribiendo en el receptor: 00:06:43	00:06:58

Comprobación de coherencia de datos: Sin verificar

Fuente: elaboración propia

Ilustración 43. Evidenciada la inserción sobre la tabla SabanaReporte.

Vista previa de los datos

Servicio vinculado: Lk_Ser_01

Objeto:

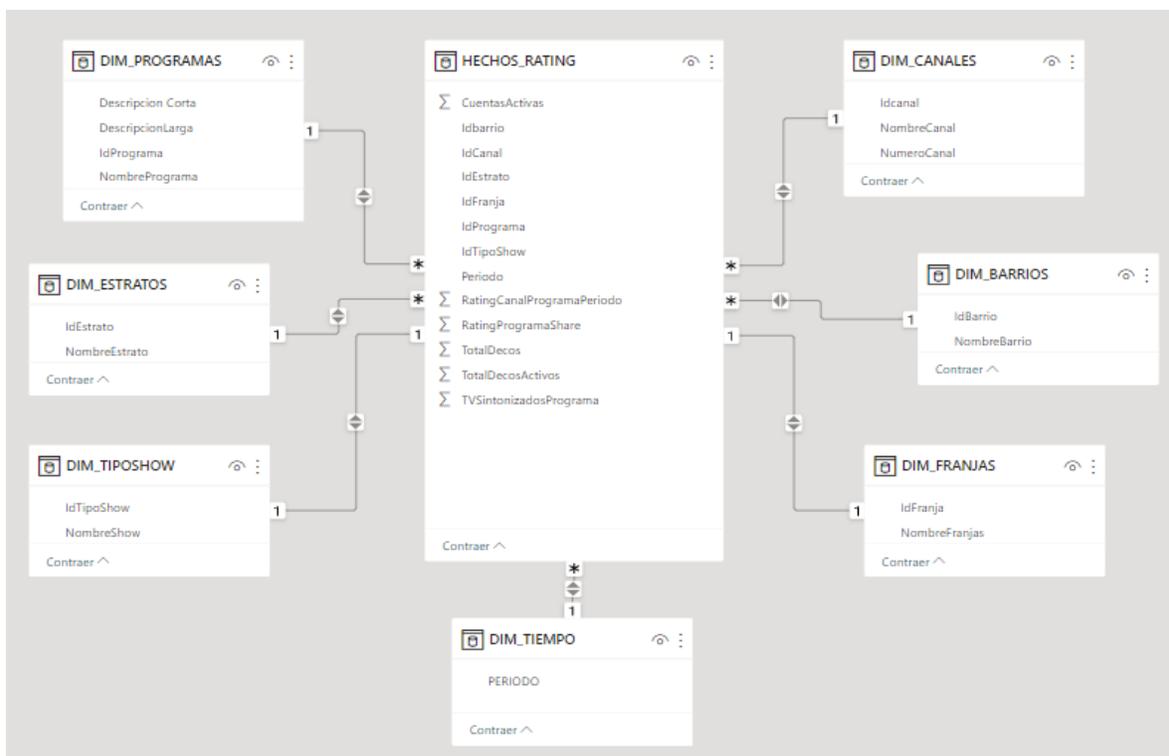
	NumeroCanalTP	NombreCanalTP	DuracionPrograma	Nombre	AnioInicioPrograma	MesInicioPrograma	DialInicioPrograma
1	728	Golden Colombia	94	Christmas Land	2022	10	28
2	734	Golden Plus Colombia	87	Christmas Land	2022	10	25
3	734	Golden Plus Colombia	87	Christmas Land	2022	10	2
4	734	Golden Plus Colombia	87	Christmas Land	2022	10	5
5	734	Golden Plus Colombia	88	Christmas Land	2022	10	6
6	734	Golden Plus Colombia	87	Christmas Land	2022	10	17
7	734	Golden Plus Colombia	87	Christmas Land	2022	10	26
8	734	Golden Plus Colombia	54	Christmas Land	2022	10	17
9	734	Golden Plus Colombia	87	Christmas Land	2022	10	3

Fuente: elaboración propia

Con la información preparada y limpia en el staging en la base **Rating_Staging_Base_04**, se da inicio al trabajo de diseño de la solución en power bi para el cálculo del indicador de Rating, Share y Ranking. Se diseña el modelo analítico denominado estrella que permite trabajar la información contenida en la base de staging a través de tablas de dimensiones y hechos. Es te modelo facilita y genera los resultados cuantitativos a través de medidas que se recalculan automáticamente dentro del dashboard según los filtros o segmentaciones seleccionadas.

El modelo estrella para este trabajo se presenta a continuación (Ver ilustración 44).

Ilustración 44. Modelo estrella para dashboard de rating.



Fuente: Elaboración propia

Primero se comienza a crear las medidas que serán usadas directamente en el cálculo de los indicadores. Usando las siguientes expresiones en DAX:

El cálculo del rating por canal y periodo, entendiéndose por periodo de la dimensión de tiempo determinada por el año y el mes en formato **“yyymm”**. Este rating calcula la audiencia general por canal en el periodo y se calcula dividiendo el total de visitas al canal o sintonización en tv a través del decodificador, en un momento dado sobre el total de decodificadores que pueden o no sintonizarse,

esto quiere decir que incluye todos los servicios activos y suspendidos. Cuando se habla de un servicio suspendido se hace referencia a los servicios que están inhabilitados por incumplimiento en el pago, aun así, se cuenta dentro del cálculo del indicador (Ver ilustración 45).

Ilustración 45. Formula en DAX cálculo del indicador

```
RatingCanalPeriodo = divide(BaseRanking[TvSintonizados],BaseRanking[TotalPlanta])
```

Fuente: Elaboración propia

Se realiza el cálculo del indicador Share, la diferencia con el anterior es que este solo incluye la planta de servicios activos. Por lo tanto, en comparación con el Rating como indicador de audiencia corriente, Share tendrá una desviación por encima del rating generalmente muy pequeña, pero esto depende de la cantidad de usuarios activos en el momento de la medición (Ver ilustración 46).

Ilustración 46. Formula en DAX calculo indicador share

```
RatingShare = BaseRanking[TvSintonizados]/BaseRanking[CuentasActivas]
```

Fuente: Elaboración propia

Se realiza el mismo proceso para el cálculo de indicadores share y rating, pero segmentando para las franjas horarias (Ver Ilustración 47, 48).

Ilustración 47. Formula rating para franja.

```
RatingCanalFranjaPeriodo = HechosFranja[VisitasFranja]/HechosFranja[TotalPlanta]
```

Fuente: Elaboración propia

Ilustración 48. Formula share para franja

```
RatingFranjaShare = HechosFranja[VisitasFranja]/HechosFranja[CuentasActivas]
```

Fuente: Elaboración propia

Se realiza el mismo proceso para cálculo de indicadores share y rating y, pero segmentando para los programas (Ver Ilustración 49, 50).

Ilustración 49. Formula DAX cálculo de rating para programas

```
RatingCanalProgramaPeriodo = HechosRankingCanalPrograma[TVSintonizadosPrograma]/
HechosRankingCanalPrograma[TotalPlanta]
```

Fuente: Elaboración propia

Ilustración 50. Formula DAX para cálculo de Share para programas

```
RatingProgramaShare = HechosRankingCanalPrograma[TVSintonizadosPrograma]/
HechosRankingCanalPrograma[CuentasActivas]
```

Fuente: Elaboración propia

También se crean medidas con el fin de establecer los rankings del top 10 con mayor rating y share, esto se realiza creando la siguiente medida (Ver Ilustración 51, 52).

Ilustración 51. Formula DAX cálculo de Ranking top 10 para rating

```
. Top 10 Rating =
! VAR TopNum = TOPN(10, VALUES(HechosRankingCanalPrograma[TotalPlanta]), [Porcentaje Rating],
DESC)
! RETURN CALCULATE([Porcentaje Rating], TopN)
```

Fuente: Elaboración propia

Ilustración 52. Formula DAX cálculo de Ranking top 10 para share

```
Top 10 Share =
VAR TopNum = TOPN(10, VALUES(HechosRankingCanalPrograma[CuentasActivas]), [PorcentajeShare],
DESC)
RETURN CALCULATE([Porcentaje Share], TopN)
```

Fuente: Elaboración propia

Una vez se compilan las medidas en power bi, se completan los datos para el cálculo completo de indicadores. (Ver ilustración 53)

Ilustración 53. Variables calculadas a nivel de tabla a través de las medidas

Periodo	TVSintonizadosPrograma	RatingCanalProgramaPeriodo	RatingProgramaShare
202301	90	0.87 %	0.88 %
202301	76	0.74 %	0.75 %
202301	27	0.26 %	0.27 %

Fuente: Elaboración propia

6 Resultado

Esta etapa se refiere a la entrega de los resultados a través del dashboard de cálculo de indicadores de rating y ranking para la ciudad de Villavicencio que es el front de la solución con la que Mayha podrá interactuar. Esto incluye la creación de los gráficos y otros contenidos para presentar los resultados de manera clara y eficiente.

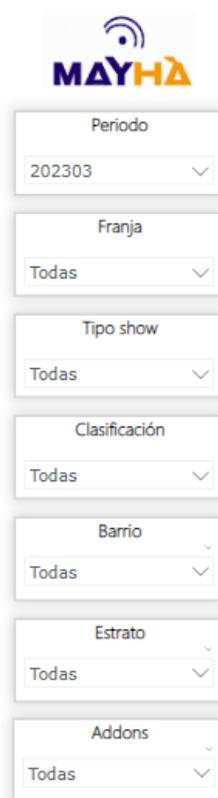
La elaboración del Dashboard se aborda realizando primero la creación de las segmentaciones o filtros que podrán generar las diferentes perspectivas en las consultas que se hagan al dashboard, segundo creando los totales de referencia que contienen las medidas generales y que se recalculan al filtrar. Por último, el desarrollo de cada una de las gráficas para que faciliten la interpretación de los valores.

Se definieron los filtros y se procedió a configurarlos en el dashboard de la siguiente forma:

- **Periodo:** dimensión de tiempo determinada por el año y el mes en formato “yyyymm”.
- **Franja:** Este filtro permite seleccionar los indicadores de una o varias franjas horarias : Early, Day, Prime, Late y Overnight explicadas anteriormente.
- **Tipo Show:** Almacena la información de clasificación del programa. El filtro permitirá ver la medición de rating y share para cada categoría de programas.

- **Clasificación:** Esta segmentación permite ver el cálculo de los indicadores sobre la clasificación de edad de los televidentes.
- **Barrio:** Esta segmentación permite ver el cálculo de los indicadores por barrio donde se encuentran instalados los decodificadores.
- **Estrato:** Esta segmentación permite ver el cálculo de los indicadores por el estrato de los televidentes.
- **Addons:** Esta segmentación permite ver el cálculo de los indicadores por los productos adicionales que tengan los televidentes. (Ver ilustración 54)

Ilustración 54. Relación de filtros creados en el dashboard.



MAYHA

Periodo
202303

Franja
Todas

Tipo show
Todas

Clasificación
Todas

Barrio
Todas

Estrato
Todas

Addons
Todas

Fuente: Elaboración propia

Una vez diseñadas y creadas las segmentaciones se procede con el cálculo de los totales que representan los valores generales del contexto de la información. (Ver ilustración 55)

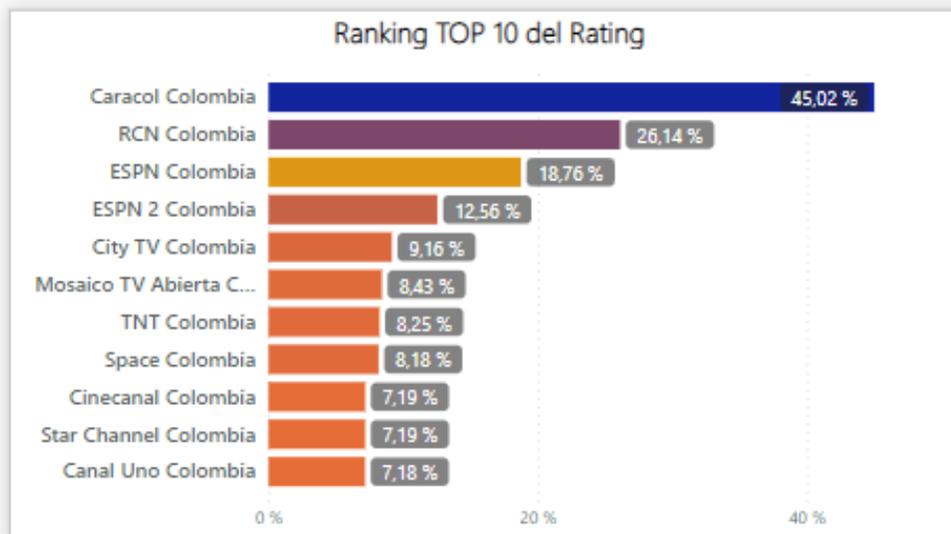
Ilustración 55. Totales de contexto general.



Fuente: Elaboración propia

El siguiente paso es comenzar a crear grafica por grafica en el siguiente orden y que permiten facilitar la interpretación de las cifras, el primer grafico es el relacionado con el top 10 del rating para dar visibilidad de primer plano a los canales con más éxito de sintonía en Mayha Comunicaciones. (Ver ilustración 56)

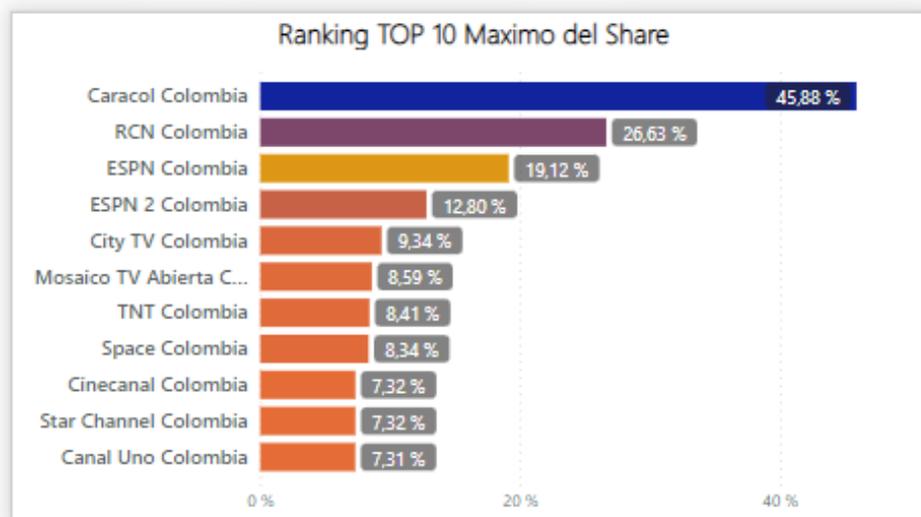
Ilustración 56. Ranking del Top 10 de Rating Máximo



Fuente: Elaboración propia

El siguiente grafico que se desarrolla es el del top 10 máximo, pero con el indicador de Share o cuota de pantalla. (Ver ilustración 57)

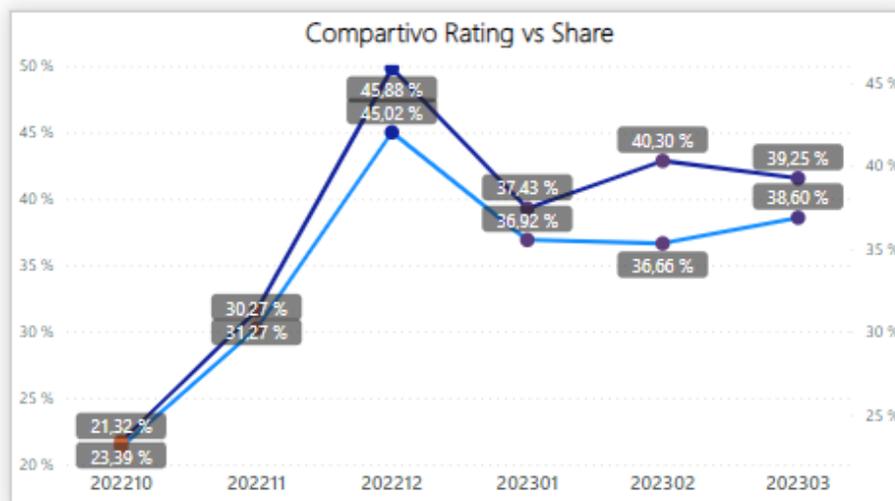
Illustration 57. Ranking del top 10 del Share.



Fuente: Elaboración propia

Con los top 10 haciendo énfasis en el éxito de los canales medidos por los dos indicadores, se diseña como tercera pieza de información un gráfico que permite evidenciar el comparativo entre los dos indicadores: Rating y Share (Ver Ilustración 58).

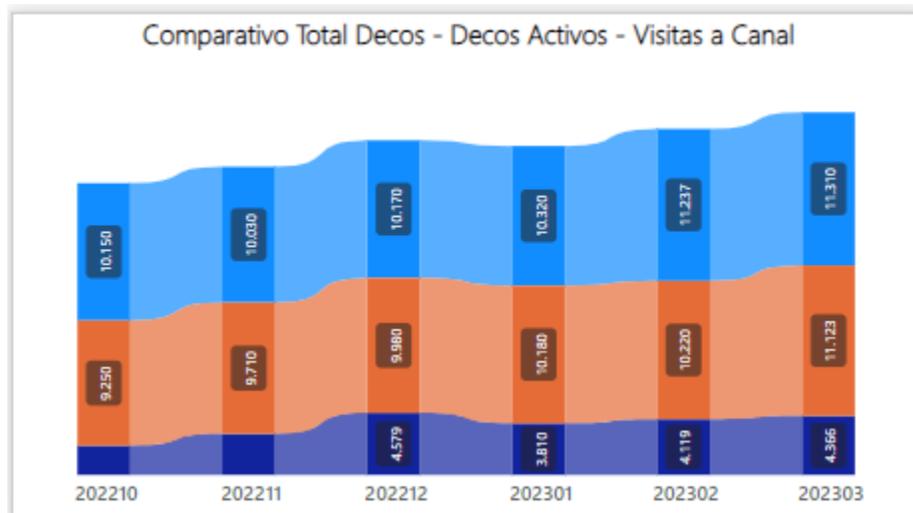
Ilustración 58. Comparativa indicadores por periodo.



Fuente: Elaboración propia

También se diseña un gráfico de tendencia con el fin de brindar información sobre las cantidades por cada periodo en cuanto a visitas a canal (sintonizaciones), total de planta por periodo y total de decodificadores activos (interactúa con la cinta de totales antes mencionada) (Ver Ilustración 59).

Ilustración 59. Comparativo de totales generales mes a mes.



Fuente: Elaboración propia

En una segunda pestaña se aborda el detalle de los indicadores por programa, esta información es bastante extensa para ser vista solo con gráfico, por lo tanto, se organiza en una matriz de datos para facilitar la visualización y consulta (Ver Ilustración 60).

Ilustración 60. Matriz periodo , programas vs indicadores.

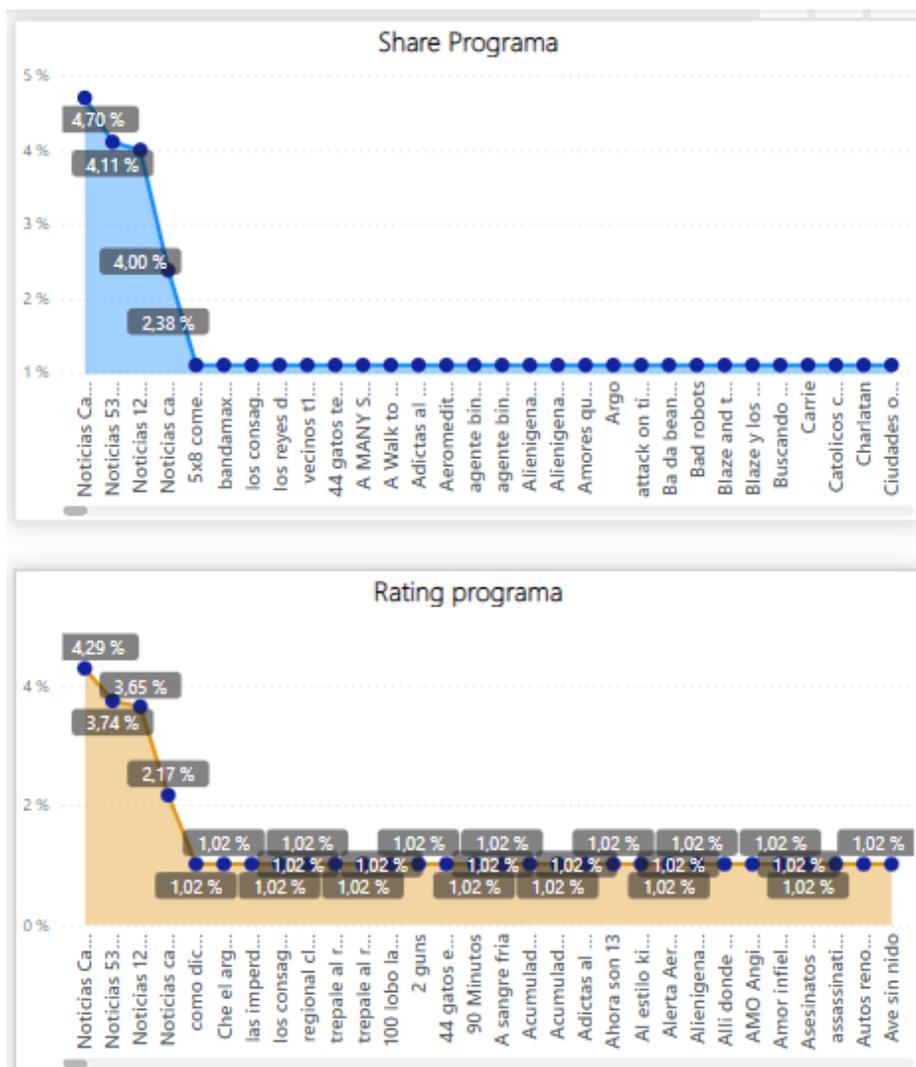
Periodo NombreCanalTP	202210		202211		202212		202301		202302	
	Rating Programa	Share Programa								
Caracol Colombia	4,29 %	4,70 %	1,00 %	1,03 %	0,99 %	1,01 %	0,99 %	1,00 %	0,91 %	1,00 %
A&E Colombia	1,00 %	1,10 %	1,02 %	1,05 %	1,00 %	1,02 %	0,99 %	1,00 %	0,90 %	0,99 %
A3 Series Colombia	0,97 %	1,06 %	1,02 %	1,05 %	0,93 %	0,95 %	0,99 %	1,00 %	0,91 %	1,00 %
AMC Colombia	1,00 %	1,10 %	1,02 %	1,05 %	1,00 %	1,02 %	0,99 %	1,00 %	0,91 %	1,00 %
Animal Planet Colombia	1,00 %	1,10 %	1,02 %	1,05 %	1,00 %	1,02 %	0,99 %	1,00 %	0,91 %	1,00 %
AXN Colombia	1,00 %	1,10 %	1,02 %	1,05 %	1,00 %	1,02 %	0,99 %	1,00 %	0,91 %	1,00 %
AZ Cinema Colombia	0,96 %	1,05 %	1,02 %	1,05 %	1,00 %	1,02 %	0,99 %	1,00 %	0,91 %	1,00 %
Bandamax Colombia	1,00 %	1,10 %	1,02 %	1,05 %	1,00 %	1,02 %	0,99 %	1,00 %		
Bitme Colombia	1,00 %	1,10 %	1,02 %	1,05 %	0,99 %	1,01 %	0,99 %	1,00 %		
Canal Capital Colombia	1,00 %	1,10 %	1,02 %	1,05 %	0,96 %	0,98 %	0,99 %	1,00 %	0,91 %	1,00 %
Canal Congreso Colombia	1,00 %	1,10 %	1,02 %	1,05 %	1,00 %	1,02 %	0,99 %	1,00 %	0,91 %	1,00 %
Canal TRO Colombia	1,00 %	1,10 %	1,02 %	1,05 %	1,00 %	1,02 %	0,99 %	1,00 %	0,91 %	1,00 %
Canal Uno Colombia	0,99 %	1,08 %	1,02 %	1,05 %	1,00 %	1,02 %	0,99 %	1,00 %	0,91 %	1,00 %
Cartoon Network Colombia	1,00 %	1,10 %	1,02 %	1,05 %	1,00 %	1,02 %	0,99 %	1,00 %	0,91 %	1,00 %

Fuente: Elaboración propia

Se relacionan dos gráficos de área bajo la curva para mostrar los picos del rating y del share por programa, también este presenta una visual sobre los programas con

mediciones estándar y las mediciones más bajas, es decir programas que no están teniendo éxito (Ver Ilustración 61).

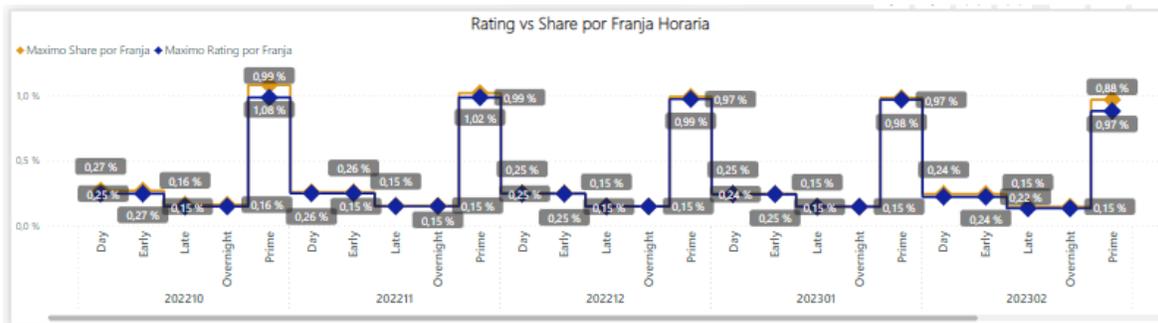
Ilustración 61. Indicadores por programa.



Fuente: Elaboración propia

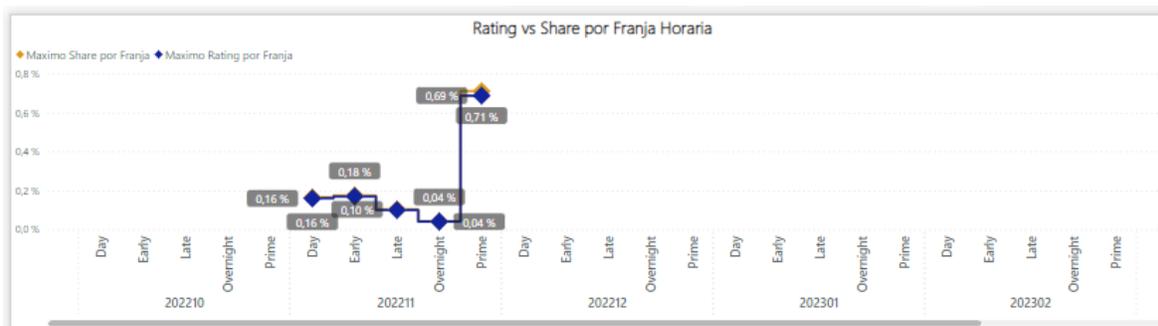
En una tercera hoja se trabaja en detalle sobre el cálculo de los indicadores basado en la franja horaria, por lo tanto, se realiza un gráfico completo que permite ser explorado para dar conocimiento de cada detalle de la medida (Ver Ilustración 62, 63).

Ilustración 62. Gráfico de los indicadores por franjas horarias.



Fuente: Elaboración propia

Ilustración 63. Gráfico de indicadores por franja horaria con filtros aplicados



Fuente: Elaboración propia

Por último, se añade el reporte de detalle de los indicadores por franjas horarias, que sirve para visualizar la información a través del periodo y hallar las medidas del cálculo del indicador de Rating Rankeado por el periodo y las franjas horarias; day, early, late, overnight y prime (Ver Ilustración 63).

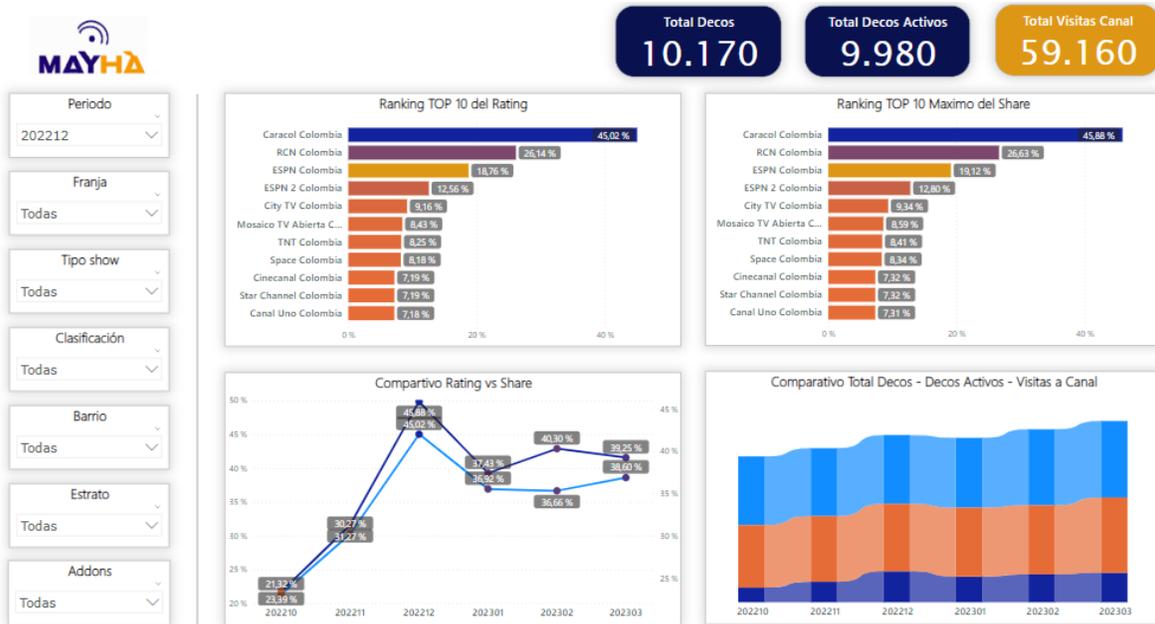
Ilustración 64. Reporte detallado distribución indicadores y por periodo y franja horaria.

Franja Nombre/Canal/TP	Day		Early		Late		Overnight		Prime		Total
	Rating por Franja	Share por Franja									
A&E Colombia	0.23%	0.23%	0.22%	0.24%	0.12%	0.12%	0.10%	0.10%	0.83%	0.84%	0.83%
202210	0.13%	0.14%	0.22%	0.24%	0.08%	0.09%	0.06%	0.06%	0.63%	0.69%	0.63%
202211	0.16%	0.16%	0.17%	0.18%	0.10%	0.10%	0.04%	0.04%	0.69%	0.71%	0.69%
202212	0.23%	0.23%	0.19%	0.19%	0.12%	0.12%	0.05%	0.05%	0.83%	0.84%	0.83%
202301	0.19%	0.20%	0.21%	0.22%	0.11%	0.11%	0.02%	0.02%	0.61%	0.62%	0.61%
202302	0.20%	0.22%	0.15%	0.17%	0.09%	0.10%	0.07%	0.08%	0.77%	0.84%	0.77%
202303	0.19%	0.19%	0.21%	0.22%	0.09%	0.09%	0.10%	0.10%	0.57%	0.58%	0.57%
A3 Series Colombia	0.18%	0.19%	0.23%	0.23%	0.14%	0.14%	0.14%	0.14%	0.99%	1.02%	0.99%
Adrenalina Sports Colombia	0.22%	0.22%	0.25%	0.27%	0.13%	0.15%	0.15%	0.16%	0.97%	1.06%	0.97%
AMC Colombia	0.24%	0.25%	0.21%	0.21%	0.10%	0.10%	0.14%	0.15%	0.92%	1.01%	0.92%
Animal Planet Colombia	0.21%	0.23%	0.25%	0.26%	0.15%	0.16%	0.13%	0.13%	0.76%	0.78%	0.76%
Antena 3 Colombia	0.19%	0.19%	0.25%	0.25%	0.15%	0.16%	0.14%	0.14%	0.85%	0.94%	0.85%
AXN Colombia	0.25%	0.26%	0.25%	0.25%	0.15%	0.16%	0.13%	0.13%	0.98%	1.01%	0.98%
AZ Cinema Colombia	0.24%	0.25%	0.23%	0.25%	0.14%	0.14%	0.12%	0.13%	0.80%	0.88%	0.80%
Az Clic Colombia	0.23%	0.25%	0.23%	0.23%	0.13%	0.14%	0.13%	0.13%	0.95%	1.04%	0.95%
AZ Mundo Colombia	0.25%	0.25%	0.22%	0.22%	0.13%	0.13%	0.15%	0.15%	0.82%	0.84%	0.82%
Total	0.25%	0.27%	0.25%	0.27%	0.15%	0.16%	0.15%	0.16%	0.99%	1.08%	0.99%

Fuente: Elaboración propia

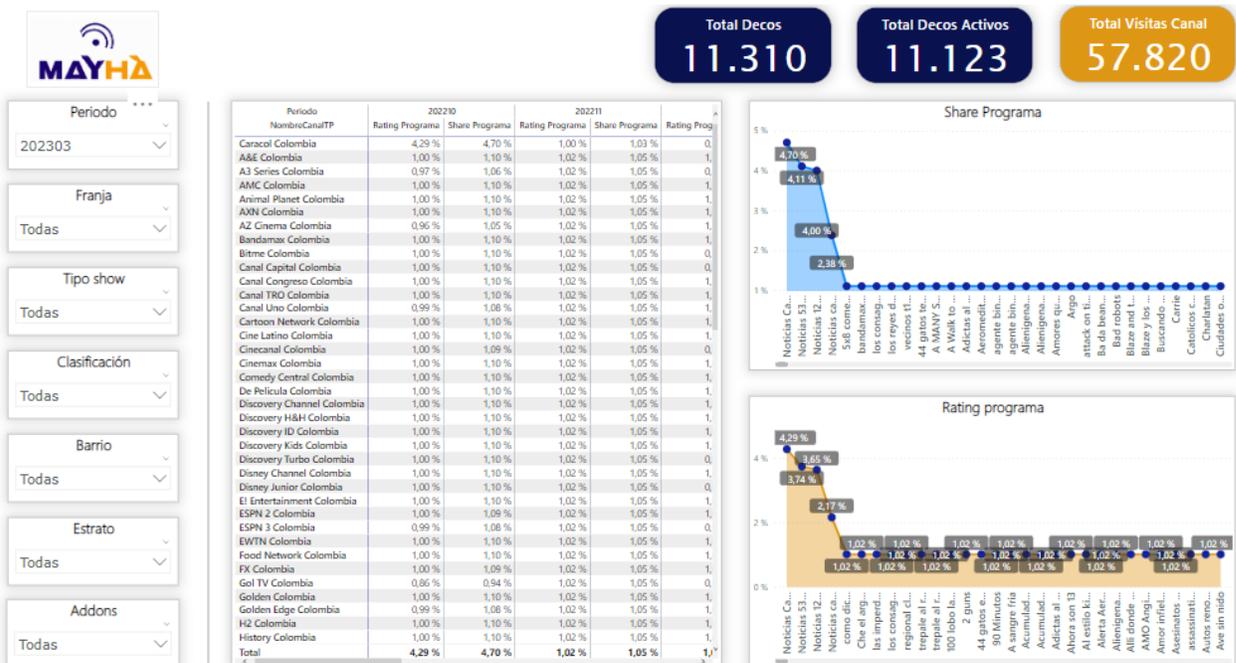
Estas son las vistas de cada página del dashboard (Ver Ilustración 65, 66, 67):

Ilustración 65. Dashboard Página 1



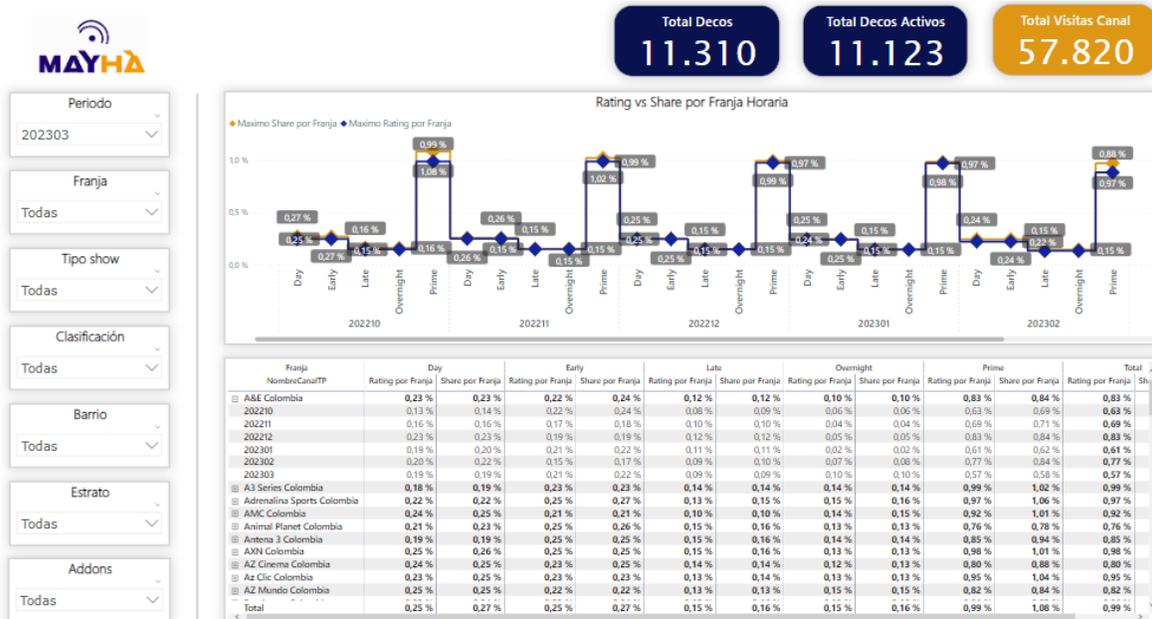
Fuente: Elaboración propia

Ilustración 66. Dashboard Página 2



Fuente: Elaboración propia

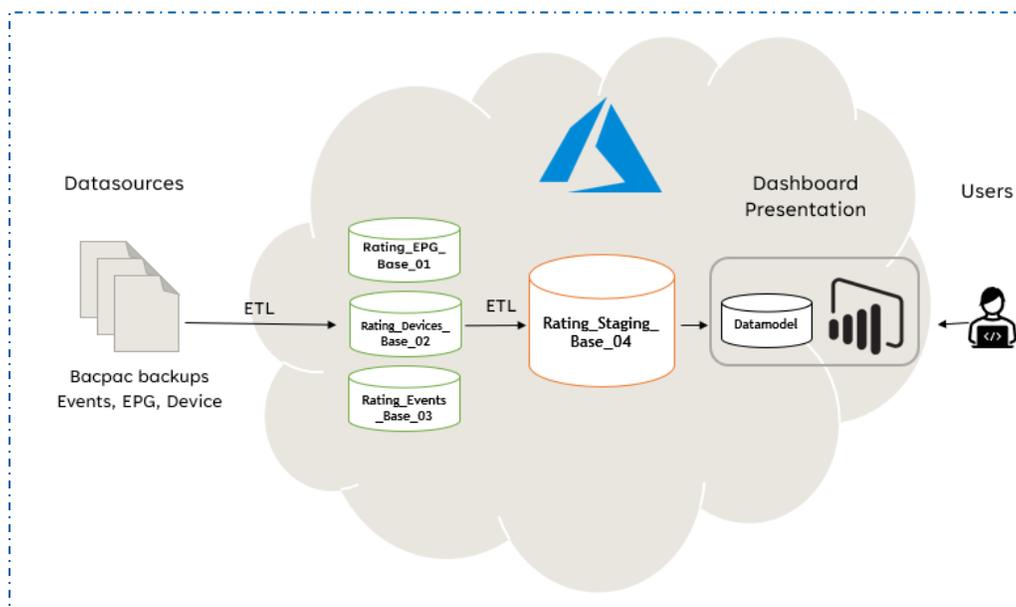
Ilustración 67. Dashboard Página 3



Fuente: Elaboración propia

Por último, se representa la arquitectura implementada para este trabajo (Ver Ilustración 68)

Ilustración 68. Arquitectura implementada.



Fuente: Elaboración propia

7 Conclusiones

Al culminar un proyecto, es importante revisar los resultados obtenidos y analizar si se lograron cumplir los objetivos planteados. Para ello, se pueden identificar una serie de conclusiones que permitan sintetizar lo que se logró en el proyecto y evaluar su impacto. En este sentido, en el siguiente listado se presentan las conclusiones, el cual permitió obtener información relevante sobre los canales de televisión en Villavicencio y diseñar un dashboard que facilita la interpretación y análisis de los datos.

1. Se logró identificar las fuentes de datos relevantes necesarias para el cálculo del rating y ranking de los canales de televisión en Villavicencio.
2. Se realizaron los procesos de Extracción, Transformación y Carga (ETL) necesarios para obtener una data consistente para el análisis de la información referente a los canales de televisión en Villavicencio.
3. Se obtuvo una data consistente que permitió realizar análisis y visualizaciones precisas de la información sobre los canales de televisión en Villavicencio.
4. Se diseñaron las características funcionales del dashboard, lo que permitió visualizar la información sobre los canales de televisión de manera eficiente y fácil de entender.
5. También se diseñaron las características no funcionales del dashboard, asegurando que el dashboard cumpliera con los requisitos de desempeño, seguridad y escalabilidad.
6. Se logró implementar el dashboard que muestra el rating y ranking de los canales de televisión en Villavicencio.
7. El dashboard permite visualizar los datos de forma clara y sencilla, facilitando la interpretación y análisis de los mismos.
8. La implementación del dashboard con el rating y ranking de los canales de televisión en Villavicencio permite a los usuarios tomar decisiones basadas en datos precisos y confiables.

8 Referencias

- Richardson (2017), L. and Ruby, S. and Hansson, D.H. RESTful Web Services. North Carolina: O'Reilly Media, Incorporated.
- Gómez F. (2013), Notas del curso, Bases de datos. México, D.F.: Casa Abierta - Universidad Autónoma de México.
- Curto D. (2012) Introducción al Business Intelligence. Barcelona: Editorial UOC
- Sandell M. (2020) Visualización de Datos & Storytelling. New York: Editorial Covarrubias.
- Gómez, J. (2017). Comunicación audiovisual: conceptos básicos y producción de contenidos. Madrid: Síntesis.
- Martínez, J. (2019). Análisis de datos de audiencia en canales de televisión. Bogotá: Editorial Universitaria.
- Martínez, J. (2019). Análisis de datos en televisión: estrategias y técnicas. Bogotá: Universidad de los Andes.
- García, L. (2020). Big data en televisión: técnicas de gestión y análisis. Barcelona: UOC.
- Ramírez, J. (2018). Bases de datos: conceptos y aplicaciones. México: Pearson.
- Pérez, A. (2019). Fundamentos de los gestores de bases de datos. Madrid: RA-MA Editorial.
- Cavaller, V. (2014). Análisis de audiencias y estrategias de visibilidad. España: Editorial UOC.
- Sánchez, A. (2019). Análisis de datos y visualización en canales de televisión. Bogotá: Editorial Universitaria.
- Real Academia Española. (s.f.). Canal. En Diccionario de la lengua española. Recuperado el 27 de septiembre de 2021, de <https://dle.rae.es/canal>
- Pérez, J. (2022). Decodificador. En Diccionario de Telecomunicaciones. Recuperado el 18 de marzo de 2023, de <https://www.diccionariotelecomunicaciones.es/decodificador/>
- Pérez, J. (2020). Medición de audiencia televisiva. Ediciones de la Universidad de Colombia.
- Ruiz, J. M. (2016). Comunicación Audiovisual. Editorial Síntesis.
- Ministerio de Tecnologías de la Información y las Comunicaciones (s.f.). ¿Qué es el MINTIC? https://www.mintic.gov.co/portal/inicio/MINTIC/Que_es_el_MINTIC/

- Kantar IBOPE Media Colombia. (s.f.). Metodología PeopleMeter. Recuperado el 25 de febrero de 2023, de https://www.kantaribopemedia.com.co/audiencia_tv.html
- Nielsen Colombia. (s.f.). Nuestra metodología. Recuperado el 25 de febrero de 2023, de https://www.kantaribopemedia.com.co/audiencia_tv.html
- Dircomfidencial. (2020, mayo 28). ¿Qué es la audiencia media? Dircomfidencial. <https://dircomfidencial.com/diccionario/que-es-la-audiencia-media-20200528-1237/>
- ComScore. (s.f.). Television measurement. Recuperado el 25 de febrero de 2023, de <https://www.comscore.com/lat/Prensa-y-Eventos/Blog/Ranking-de-Medios-auditados-por-Comscore-en-Colombia>
- Asociación Colombiana de Investigación de Medios. (2021). Estudio General de Medios. Recuperado el 25 de febrero de 2023, de <http://acim.org.co/investigaciones/estudio-general-de-medios/>
- Banrepcultural. (s.f.). La televisión en Colombia. Enciclopedia. Recuperado el 11 de mayo de 2023, de https://enciclopedia.banrepcultural.org/index.php/La_televisi%C3%B3n_en_Colombia
- Nielsen. (s.f.). What is ratings? Recuperado el 25 de febrero de 2023, de <https://www.nielsen.com/us/en/measurement/television-measurement/ratings/>
- Talend. (s.f.). Data cleansing. Recuperado el 25 de febrero de 2023, de <https://www.talend.com/resources/data-cleansing/>
- Referencia: Escobar, Ruiz, S., Sosa, Vazquez, S., Espinal Martín, Y., & Puebla Martínez, M. (2012). Aplicación para normalizar bases de datos relacionales. En REVISTA DIGITAL SOCIEDAD de LA INFORMACIÓN, 38. <http://www.sociedadelainformación.com/38/Relacional.pdf>
- Decodificador de TV ONT GPON HUAWEI HG8245H. Recuperado de 8 Marzo de 2023, de <https://e.huawei.com/en/products/optical-terminal>

Anexo 1

Diccionario de datos base de datos EPG

TABLE	TABLE NAME	COLUMN NAME	DESCRIPTION	IS NULLABLE	DATA TYPE	LENGHT
IPTV	CategoriaPrograma	IdCategoria	Identificador Categoría	NO	int	NULL
IPTV	CategoriaPrograma	IdClasificacion	Identificador Clasificación programa	YES	int	NULL
IPTV	CategoriaPrograma	IdGenero	Identificador genero (Acción, Aventura etc)	YES	int	NULL
IPTV	CategoriaPrograma	IdTipoShow	clasificación más general del genero	YES	int	NULL
IPTV	CategoriaPrograma	Nombre	Nombre del programa	YES	varchar	130
IPTV	CategoriaPrograma	Descripcion	descripción de la categoría	YES	varchar	-1
IPTV	CategoriaPrograma	Actores	Nombres de actores más relevantes	YES	varchar	180
IPTV	CategoriaPrograma	NombreOrigen		YES	varchar	180
IPTV	CategoriaPrograma	CreateDate	Fecha de creación de la categoría	YES	date	NULL
IPTV	CategoriaPrograma	depurado	¿Ha recibido procesos de depuración ?	YES	varchar	400
IPTV	CategoriaPrograma	Enviado	Se ha actualizado últimamente en el deco de los clientes	YES	int	NULL
IPTV	clasificación	CreateDate	Fecha de creación de la clasificación	YES	date	NULL
IPTV	clasificación	IdClasificacion	Identificador	NO	int	NULL
IPTV	clasificación	clasificación	Nombre de clasificación (Todos, +16 etc)	YES	varchar	10

IPTV	clasificación	Descripcion	Descripción de la clasificación	YES	varchar	255
IPTV	Genero	CreateDate	Fecha de creación del genero	YES	date	NULL
IPTV	Genero	IdGenero	Identificador	NO	int	NULL
IPTV	Genero	Genero	Nombre del genero	YES	varchar	255
IPTV	Idioma	CreateDate	Fecha creación idioma	YES	date	NULL
IPTV	Idioma	IdIdioma	Identificador	NO	int	NULL
IPTV	Idioma	Idioma	Nombre del idioma	YES	varchar	255
IPTV	TipoShow	IdTipoShow	Identificador	NO	int	NULL
IPTV	TipoShow	CreateDate	clasificación del tipo show	YES	date	NULL
IPTV	TipoShow	TipoShow	Nombre del tipo show, clasificación agrupadora del genero	YES	varchar	255
IPTV	Canal	IdCanalTP	Identificador	NO	bigint	NULL
IPTV	Canal	NumeroCanalTP	Numero de canal	YES	int	NULL
IPTV	Canal	NombreCanalTP	Nombre de canal	YES	varchar	100
IPTV	Programas	IdPrograma	Identificador	NO	bigint	NULL
IPTV	Programas	IdCategoria	Identificador Categoria	YES	int	NULL
IPTV	Programas	IdClasificacion	Identificador Clasificación programa	YES	int	NULL
IPTV	Programas	IdIdioma	Identificador idioma	YES	int	NULL
IPTV	Programas	IdTipoShow	Identificador de tipo show	YES	int	NULL
IPTV	Programas	IdGenero	identificador de genero	YES	int	NULL
IPTV	Programas	Nombre	Nombre del programa	YES	nvarchar	180
IPTV	Programas	Descripcion	Descripción larga del programa	YES	nvarchar	-1
IPTV	Programas	DescripcionCorta	descripción corta del programa	YES	nvarchar	-1

IPTV	Programas	Actores	Nombre de los actores principales del programa	YES	nvarchar	250
IPTV	Programas	NombreOrigen	Nombre original del programa	YES	nvarchar	200
IPTV	Programas	Temporada	si es serie número de temporada	YES	nvarchar	50
IPTV	Programas	Capitulo	numero de episodio	YES	nvarchar	50
IPTV	Programas	Anno	Año programa	YES	int	NULL
IPTV	Programas	duración	Duración programa	YES	numeric	NULL
IPTV	Programas	CreateDate	fecha de creación del programa en el sistema	YES	date	NULL
IPTV	Programas	Depurado	¿Ha recibido procesos de depuración ?	YES	nvarchar	255
IPTV	Programas	EpName	Nombre del episodio	YES	nvarchar	180
IPTV	Programas	Enviado	se ha enviado a los dispositivos del cliente	YES	int	NULL
IPTV	Programas	DepuradoCat	¿Ha recibido procesos de depuración en la clasificación de la categoría?	YES	nvarchar	400
IPTV	Programas	IdProgramaViejo	Anteriormente tenía otro id	YES	int	NULL
IPTV	Programas	DepuradoViejo	¿Ha recibido procesos de depuración en la clasificación de la categoría?	YES	nvarchar	400
IPTV	Programas	FechaEnvio	Ultima fecha de envio a los dispositivos del cliente	YES	date	NULL

IPTV	Programas	Producto	paquete o plan que contiene el producto	YES	nvarchar	50
IPTV	Programas	idprogramaProveedor	Id de programadora	YES	bigint	NULL
IPTV	EPGHistorico	IdCanal	Identificador canal	YES	bigint	NULL
IPTV	EPGHistorico	IdPrograma	Identificador programa	NO	bigint	NULL
IPTV	EPGHistorico	AnioInicioPrograma	Anio inicio programa	YES	int	NULL
IPTV	EPGHistorico	MesInicioPrograma	Mes inicio programa	YES	int	NULL
IPTV	EPGHistorico	DiaInicioPrograma	Dia inicio programa	YES	int	NULL
IPTV	EPGHistorico	HoraInicioPrograma	Hora inicio programa	YES	int	NULL
IPTV	EPGHistorico	MinutoInicioPrograma	Minuto inicio programa	YES	int	NULL
IPTV	EPGHistorico	AnioFinPrograma	Anio fin programa	YES	int	NULL
IPTV	EPGHistorico	MesFinPrograma	Mes fin programa	YES	int	NULL
IPTV	EPGHistorico	DiaFinPrograma	Dia fin programa	YES	int	NULL
IPTV	EPGHistorico	HoraFinPrograma	Hora fin programa	YES	int	NULL
IPTV	EPGHistorico	MinutoFinPrograma	Minuto fin programa	YES	int	NULL
IPTV	EPGHistorico	DuracionPrograma	Duracion identificado programa	YES	int	NULL

ANEXO 2

Diccionario de datos de la tabla de eventos

TABLE SCHEMA	TABLE NAME	COLUMN NAME	DESCRIPTION	IS NULLABLE	DATA TYPE	CHARACTER MAXIMUM LENGTH
EVENTS	Event	IdEvent	Identificador de Evento	NO	int	
EVENTS	Event	TypeDevice	Tipo de dispositivo: Modem o Decodificador	NO	varchar	20
EVENTS	Event	MACDevice	Código mac del dispositivo	NO	varchar	17
EVENTS	Event	SerialDevice	Serial número único identificador del dispositivo	NO	int	
EVENTS	Event	StartNameDay	Nombre del día con el que inicia sintonización	NO	varchar	10
EVENTS	Event	StartTuningTime	Fecha con la que inicia la sintonización	NO	datetime	
EVENTS	Event	EndTuningTime	Fecha con la que finaliza la sintonización	NO	datetime	
EVENTS	Event	EndNameDay	Nombre del día en el que finaliza la sintonización	NO	varchar	10
EVENTS	Event	Duration	Tiempo total de sintonización	NO	int	
EVENTS	Event	TypeTuning	Tipo de sintonización : Surfing (Canaleo), Tunning (Sintonizado)	NO	varchar	10
EVENTS	Event	StripStart	Franja donde inicia la sintonización	NO	varchar	35
EVENTS	Event	StripStartReference	Horario inicial de referencia de la franja	YES	int	

EVENTS	Event	StripEndReference	Horario final de referencia de la franja	YES	int	
EVENTS	Event	StripEnd	Franja donde finaliza la sintonización	NO	varchar	35
EVENTS	Event	StripEndStartReference	Horario de franja final donde inicia la sintonización	YES	int	NULL
EVENTS	Event	StripEndEndReference	Horario de franja final donde finaliza la sintonización	YES	int	
EVENTS	Event	Week	Numero de semana de la finaliza la sintonización	NO	int	

ANEXO 3

Diccionario de datos de dispositivos instalados

TABLE SCHEMA	TABLE NAME	COLUMN NAME	DESCRIPTION	IS NULLABLE	DATA TYPE	CHARACTER MAXIMUM LENGTH
dbo	InventoryDevice	MACDevice	Código mac del dispositivo	YES	varchar	20
dbo	InventoryDevice	SerialDevice	Serial del dispositivo	NO	varchar	20
dbo	InventoryDevice	TypeDevice	Tipo de dispositivo: Modem o Decodificador	YES	varchar	25
dbo	InventoryDevice	CodCity	Código DANE del departamento, municipio	YES	varchar	10
dbo	InventoryDevice	Neighborhood	Nombre del Barrio	YES	varchar	60
dbo	InventoryDevice	EconomicCode	Estrato	YES	int	NULL

ANEXO 4

Directiva para Entrega de insumos



DIRECTIVA DE GOBIERNO DE DATOS DE MAYHA COMUNICACIONES

FECHA DE CREACIÓN		NOMBRE DEL LINEAMIENTO
10-03-2023		Entrega de insumos para desarrollo de Dashboard de Rating y Ranking
POLITICA NO.	VERSIÓN NO.	CREADO POR
1	1	Comité de datos Mayha

I. ENTREGABLE / FUNCIONALIDAD QUE SE PRETENDE GARANTIZAR CON LA POLITICA

Entrega de insumos con datos dummy para desarrollo de dashboard que **permita el cálculo de los indicadores de rating rankeado.**

II. PROPÓSITO DE LA POLÍTICA

Describir las condiciones y características de los datos que serán entregados como insumo para la fabricación del dashboard de medición de audiencias. Este dashboard es requerido por las diferentes áreas del negocio y debe ser desarrollado con datos dummy inicialmente contenidos en las bases de datos relacionadas a la emisión de programas de tv, servicio que presta Mayha Comunicaciones a sus clientes. En resumen, el propósito de la política es asegurar que se utilicen los datos adecuados para construir el dashboard de medición de audiencias, proteger la privacidad de los clientes y garantizar la funcionalidad estándar del producto final.

III. ALCANCE DE LA POLÍTICA

La información debe entregarse en el formato generado por el motor de base de datos que los contiene, por lo tanto, es apropiado entregar la información en .bak o .SQL dummy de la información correspondiente a 6 meses de :

1. Base de datos de Guía Electrónica de Programación cuyo nombre es **EPG**. Que está contenida en la instancia de pruebas servidor: **TestEPG**.
2. Base de datos de dispositivos instalados cuyo nombre es **Devices**. Que se encuentra almacenada en la instancia de pruebas servidor: **Test**. **NO DEBE CONTENER IDENTIFICACIONES NI NINGUN TIPO DE DATO DEL CLIENTE**
3. Tabla **EVENTS** extracto de la base de datos denominada **EventChannel**: Que se encuentra almacenada en la instancia de pruebas servidor: **test**.

IV. REGLAS DE POLÍTICA

A. ACCESO A DATOS	El área de Arquitectura tecnológica realizará la extracción de la información relacionada con el alcance de la política descrito y será la única área con acceso directo a esta fuente de datos para extraer y entregar.
B. USO DE DATOS	La información será entregada exclusivamente para la construcción del dashboard de medición de audiencia.
C. INTEGRACIÓN DE DATOS	Los datos pueden ser integrados al modelo que los desarrolladores del dashboard de audiencias requieran en el proceso del desarrollo.
D. INTEGRIDAD DE LOS DATOS	<i>Los datos almacenados a entregar son dummy, no gozan de una calidad estándar y el resultado de las mediciones del dashboard puede presentar desviaciones que son tolerables para este ejercicio.</i>

V. ROLES / RESPONSABILIDADES

ROL	RESPONSABILIDAD
Administrador de datos	Estará encargado de la coordinación para la extracción, y entrega de los archivos al ingeniero de datos que desarrollará el dashboard según la presente política.
Ingeniero de datos	Desarrollador que elabora el dashboard versión datos dummy.
Data Steward	Indica cuales son las bases de datos necesarias para el desarrollo del dashboard y verifica que la entrega se de en las condiciones acordadas en la presente política.
Chif Data Officer: CDO	Indica el alcance de la política con respecto a la necesidad de la compañía y se asegura de que se dé cumplimiento a lo acordado.

VI. PROCESO DE REVISIÓN

Una vez el **administrador de datos** hace la extracción de la información relacionada en la presente política, el **data steward** hará la verificación de que se cumpla con las especificaciones de formato, origen y tipo de información que será suministrada. Este a su vez le notificará el CDO acerca del proceso para que se dé el visto bueno y se ejecute la entrega al **ingeniero de datos** desarrollador.

VII. RECURSOS

TIPO DE MATERIAL	NOMBRE	UBICACIÓN / ENLACE / SERVER
Archivo	EPG.bacpac	TestEPG
Archivo	Devices.bacpac	Test
Archivo	Events.bacpac	Test

VIII. CONTACTOS

NOMBRE Y TÍTULO	TELÉFONO	CORREO ELECTRÓNICO
Juan García* / Administrador de datos	32022233115*	a@mayha.com*
Leonardo Marin* / Ingeniero de datos desarrollador	31552525526*	b@mayha.com*
Steward Mendoza* / Data Steward	52636363341*	c@mayha.com*
Julio Hernandez* / CDO	41414144114*	d@mayha.com*

- Los nombre propios, telefonos y correos de los funcionarios fueron cambiados

IX. DEFINICIONES / TERMINOLOGÍA

TÉRMINO	DEFINICIÓN
EPG	Base que contiene la información emitida por las programadoras a través de los dispositivos de la compañía.
Devices	Dispositivos de TV que instala Mayha a sus clientes como: Decodificadores y módems de transmisión de datos.
Rating	Indicador de audiencia

ANEXO 5

Directiva para Nomenclatura de Objetos



DIRECTIVA DE GOBIERNO DE DATOS DE MAYHA COMUNICACIONES

FECHA DE CREACIÓN		NOMBRE DE LA DIRECTIVA
10-03-2023		Nomenclatura de los objetos a crear en Microsoft Azure
POLITICA NO.	VERSIÓN NO.	CREADO POR
2	1	Comité de datos Mayha

I. ENTREGABLE / FUNCIONALIDAD QUE SE PRETENDE GARANTIZAR CON LA POLÍTICA

Marcado de objetos y servidores en la nube de Microsoft Azure para el desarrollo del dashboard de medición de rating.

II. PROPÓSITO DE LA POLÍTICA

Describir los nombres que deben llevar los objetos y servidores en la nube de Microsoft para facilitar la identificación y desarrollo del dashboard para Mayha Comunicaciones.

III. ALCANCE DE LA POLÍTICA

Los nombres de los objetos que se deben marcar son los siguientes:

Grupo de recursos: Debe nombrarse como “ProyectoUan2023”.

Servidor SQL: Debe tener el nombre de: “psserprycan”.

Base de datos EPG: En Azure debe nombrarse como: “Rating_EPG_Base_01”

Base de datos Devices: En Azure debe nombrarse como: “Rating_Devices_Base_02”

Base de datos Events: En Azure debe nombrarse como: “Rating_Events_Base_03”

Base de datos Staging: En Azure debe nombrarse como: “Rating_Staging_Base_04”

IV. REGLAS DE POLÍTICA

E. ACCESO A DATOS	Los ingenieros de datos pueden usar la información dummy suministrada y cargarla en Microsoft Azure cumpliendo con las reglas estipuladas en esta política.
F. USO DE DATOS	La información será entregada exclusivamente para la construcción del dashboard de medición de audiencia.
G. INTEGRACIÓN DE DATOS	Los datos pueden ser integrados al modelo que los desarrolladores del dashboard de audiencias requieran en el proceso del desarrollo.
H. INTEGRIDAD DE LOS DATOS	Los datos almacenados a entregar son dummy, no gozan de una calidad estándar y el resultado de las mediciones del dashboard puede presentar desviaciones que son tolerables para este ejercicio.

V. ROLES / RESPONSABILIDADES

ROL	RESPONSABILIDAD
Ingeniero de datos	Desarrollador que elabora el dashboard versión datos dummy.
Data Steward	Verificará que las bases de datos creadas en el sitio suministrado para el desarrollo cumplan con los nombres descritos en esta política.
Chif Data Officer: CDO	Indica el alcance de la política con respecto a la necesidad de la compañía y se asegura de que se de cumplimiento a lo acordado.

VI. PROCESO DE REVISIÓN

Una vez se posean los datos entregados por el administrador de base de datos se debe realizar el marcado de las bases de datos en Microsoft Azure como está indicado en la presente política, el data steward hará la verificación de que se cumpla con las especificaciones y a su vez le notificara el CDO acerca del proceso para que se dé el visto bueno para que se dé inicio a los desarrollos solicitados.

VII. RECURSOS

TIPO DE MATERIAL	NOMBRE	UBICACIÓN / ENLACE / SERVER
Archivo	Plantilla política nro. 2	

VIII. CONTACTOS

NOMBRE Y TÍTULO	TELÉFONO	CORREO ELECTRÓNICO
-----------------	----------	--------------------

Juan García* / Administrador de datos	32022233115*	a@mayha.com*
Leonardo Marin* / Ingeniero de datos desarrollador	31552525526*	b@mayha.com*
Steward Mendoza* / Data Steward	52636363341*	c@mayha.com*
Julio Hernandez* / CDO	41414144114*	d@mayha.com*

- Los nombre propios, telefonos y correos de los funcionarios fueron cambiados

IX. DEFINICIONES / TERMINOLOGÍA

TÉRMINO	DEFINICIÓN
Marcación de objetos	Nombre que se le dará a los servidores y bases de datos en el servicio de Microsoft Azure en la nube.
Devices	Dispositivos de TV que instala Mayha a sus clientes como: Decodificadores y módems de transmisión de datos.
Rating	Indicador de audiencia

ANEXO 6

Directiva para Reglas de negocio



DIRECTIVA DE GOBIERNO DE DATOS DE MAYHA COMUNICACIONES

FECHA DE CREACIÓN		NOMBRE DE LA DIRECTIVA
10-03-2023		Reglas de negocio a aplicar y cargue de datos limpios a la base "Rating_Staging_Base_04"
POLITICA NO.	VERSIÓN NO.	CREADO POR
3	1	Comité de datos Mayha

1. ENTREGABLE / FUNCIONALIDAD QUE SE PRETENDE GARANTIZAR CON LA POLÍTICA

Información verificada y limpia que cumple con las reglas determinadas en la presente política.

2. PROPÓSITO DE LA POLÍTICA

Establecer reglas sobre los datos de TV las cuales son para garantizar una correcta clasificación y medición en el dashboard.

3. ALCANCE DE LA POLÍTICA

1. De la base de datos **Rating_EPG_Base_01**, no deben tenerse en cuenta para las mediciones del rating todos los programas que en ***IPTV.Programas.Nombre*** contenga un nombre de programa "Infomercial", "Infomerciales". En los meses de octubre, noviembre, diciembre, enero, febrero y marzo. Entre los años 2022 y 2023
2. De la base de datos **Rating_EPG_Base_01**, no deben tenerse en cuenta para las mediciones del rating todos los programas cuya duración sea menor o igual 5 minutos.
3. Para la base de datos **Rating_Devices_Base_03** que contiene en la tabla de Events la columna "TypeTunning" los datos que debe usarse para las mediciones de rating es "Tunning" y para las mediciones de Zapping se debe usar únicamente "Surfing".

4. REGLAS DE POLÍTICA

I. ACCESO A DATOS	Los ingenieros de datos pueden usar la información dummy suministrada y cargarla en Microsoft Azure realizando el filtrado a través de scripts, procedimientos almacenados o ETL que aplique las anteriores reglas descritas en esta política.
J. USO DE DATOS	La información será entregada exclusivamente para la construcción del dashboard de medición de audiencia.
K. INTEGRACIÓN DE DATOS	Los datos pueden ser integrados al modelo que los desarrolladores del dashboard de audiencias requieran en el proceso del desarrollo.
L. INTEGRIDAD DE LOS DATOS	Los datos almacenados a entregar son dummy, no gozan de una calidad estándar y el resultado de las mediciones del dashboard puede presentar desviaciones que son tolerables para este ejercicio.

5. ROLES / RESPONSABILIDADES

ROL	RESPONSABILIDAD
Ingeniero de datos	Desarrollador que elabora el dashboard versión datos dummy.
Ingenieros de QA	Verificara que las bases de datos cumplan con las reglas descritas en esta política.
Chif Data Officer: CDO	Indica el alcance de la política con respecto a la necesidad de la compañía y se asegura de que se de cumplimiento a lo acordado.

6. PROCESO DE REVISIÓN

Los ingenieros de datos pueden usar la información dummy suministrada y cargarla en Microsoft Azure realizando el filtrado a través de scripts, procedimientos almacenados o ETL que aplique las anteriores reglas descritas en esta política. Los ingenieros de QA verificaran en los datos que estas reglas estén aplicadas para garantizar una mayor consistencia en los datos a pesar de tratarse de información de prueba. Los ingenieros QA notificaran al CDO sobre el éxito de la aplicación de las reglas.

7. RECURSOS

TIPO DE MATERIAL	NOMBRE	UBICACIÓN / ENLACE / SERVER
Archivo	Plantilla política nro. 3	

8. CONTACTOS

NOMBRE Y TÍTULO	TELÉFONO	CORREO ELECTRÓNICO
-----------------	----------	--------------------

Juan García* / Administrador de datos	32022233115*	a@mayha.com*
Leonardo Marin* / Ingeniero de datos desarrollador	31552525526*	b@mayha.com*
Steward Mendoza* / Data Steward	52636363341*	c@mayha.com*
Julio Hernandez* / CDO	41414144114*	d@mayha.com*

- Los nombre propios, telefonos y correos de los funcionarios fueron cambiados

9. DEFINICIONES / TERMINOLOGÍA

TÉRMINO	DEFINICIÓN
Preparación de los datos.	Reglas que deben cumplirse para asegurar una mejor información en el proceso de desarrollo.