Internship Report
Submitted for the degree of
Data Engineer

# Data Warehouse for government reporting management

Jorge Antonio Aqué González

*1. External Supervisor* Octavio Vallejo
Software engineering department
Blue Ocean Technologies

*2. Internal Supervisor* Luis Gerardo Camara Salinas
Data Engineering
Universidad Politécnica de Yucatán

April 29, 2022

# Contents

Acknowledgments v

Abstract vii

1 Introduction 1
1.1 Background 1
1.2 Problem Statement 1
1.3 Justification 2
1.4 Scope and limitations 2
1.5 Objectives 4
1.5.1 General Objectives 4
1.5.2 Specific Objectives 4

2 Theoretical framework 5
2.1 Microsoft SQL Server Managment Studio 5
2.2 Data Warehouse 5
2.3 OLAP CUBE 5
2.4 Store Procedure 5
2.5 ETL 6
2.6 Pipeline 6
2.7 SIGG 6
2.8 EGOB 6

3 Methodology and development 7
3.1 Data collection 8
3.2 Information processing 10

4 Results 19

5 Conclusions and recommendations 21
5.1 Conclusion 21

Bibliography 22

# List of Figures

3.1 Pipeline . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
3.2 General Process . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
3.3 Analytical report P1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
3.4 Analytical report P2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10
3.5 Snowflake Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
3.6 On-demand Service . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
3.7 Star Model V1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
3.8 Firs Test . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
3.9 Second Test . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13
3.10 Saldo Final Operation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13
3.11 Star Model V2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
3.12 Star Model V2 with SP . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
3.13 Constellation model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
3.14 Table test - Hechos.MayorAuxiliarCuenta . . . . . . . . . . . . . . . . . . . . 16
3.15 Table test - Hechos.MayorAuxiliarPolizas . . . . . . . . . . . . . . . . . . . . 17

# Acknowledgments

I would like to thank my business advisors Octavio Vallejo and Didier Moreno since they were the ones who guided me throughout this project, they were very patient with me and thanks to them I was able to further develop my analytical thinking as a future software engineer.

# Abstract

The world of data engineering can vary in complexity since there are usually projects of all kinds, during this project, it will be shown a little about how the analysis is applied for the proposal of a functional cube aimed at government reports, it is shown as It was the evolution of the whole process and how a proposal was reached which would serve for future work of the company. This project has several limitations since the data engineering area is new and in this case, as a resident of the company, I apply a base for the emergence of new projects for this area, in the same way the communication in this project is of the utmost importance since it is necessary to be constantly learning about the tools provided by the company, in addition to an exhaustive analysis of the report that is taken as a base example for solving the problem.

# 1 Introduction

Data engineering is a branch of computer science that is a profession that has recently emerged in the labor field. This profession normally causes many doubts to people since it is related to other work professions such as data analyst, data scientist and business intelligence analyst, but it is worth mentioning that all these professions emerged with the purpose of managing and analyzing data only that each one performs in different ways, in this case we are going to focus on data engineering which we could say that it is the starting point for all the other professions mentioned above and that it is also the role that will be played during this project process. On the other hand, we will see the need to apply techniques and different types of architectures for the development of data flow structures from their collection to their storage in specialized sites for their continuous and efficient use.

## 1.1 Background

Data Engineering is a scarce profession in the labor field since at the moment there are not enough people specialized in this same work. In previous years, different engineers related to the field of computing had the role of data engineers but it was not for what they had specialized in their years of study, the position of data engineer emerged as a necessity in the field of administration and data analysis. In approximately 1962, John W. Tukey coined the term data science for the first time as a definition of data management and analysis with different mathematical techniques, years later variants of different levels would emerge such as data engineering which is responsible for the acquisition , storage, transformation and management of data.

Data engineering has a high execution complexity since it has to be constantly updated since it has to handle a large number of data management tools and apply different types of skills and discipline. Currently, many companies have begun to apply data engineering techniques since the purpose of this branch is the optimization of the flow for data analysis, which are very useful for companies to use in decision-making in the labor market. from start-ups to the same areas of government.

## 1.2 Problem Statement

Blue Ocean Technologies is a company specialized in the integration and development of software solutions which has a specialized area in the management of government data; Within this area, the management of government reports is vital for decision making for both the company and clients, therefore, as a basis, there is a flow of data for the creation of reports that ranges from the collection of data that is through from its own internal platform to the visualization of processed data on a report according to its nature. This process can normally be inefficient in waiting times since, depending on the amount of data that needs to be analyzed, more time and resources will take the generation of reports, therefore, this project focuses on the proposal of an alternative to the generation of reports. reports for the solution of reduction in waiting times. In general, the creation of data warehouses and olap cubes was proposed as the alternative to reduce waiting times, the reason being that the architecture of these data engineering techniques allows us to facilitate the way of managing and analyzing data since they are database dedicated to specific solutions.

Currently, customer reports are handled on the SIGG platform, which requires advanced knowledge of its management, since the information to be displayed in them requires prior treatment based on the understanding of the relationship between system tables, configurations that vary the interpretation of the information and particular knowledge of the technology used for the issuance of the reports, for which the clients depend on operations personnel or the EGOB FWS for the construction of specific reports.

In the same way, all this flow of information is handled through a main database, this transactional database is the data consumption for the reports, so it can be affected in terms of performance since there are times when that consumer demand may be greater. There are reports generated through the platform that make use of Store Procedures and Functions for filtering and calculating the information required for the report, which is on demand, therefore, it usually takes a considerable amount of time when generating certain reports. specific. There is a previous proposal as a solution, but it is not conclusive and the project was suspended at the time.

## 1.3 Justification

Work has been done on a proposed solution to the problem which aims to reduce waiting times for generating reports, therefore, the creation of a Data Warehouse dedicated to reports which will be processed through ETL has been proposed. 's and stored in OLAP cubes that will be prepared so that the client can make simple filters on simple queries and that they can use the reporting system they like. Each cube stored in the data warehouse will be periodically updated with the transactional database through the ETL's. The "Mayor Auxiliar de la cuenta" has been taken as a report, which is intended to be the example for the proposal of the restructuring of the entire flow of information of the selected reports.

## 1.4 Scope and limitations

During the beginning of the project, the limitations were made known immediately, since there are a large number of reports and areas directed by product leaders, so they have to be in constant communication to collect information on the processes, calculations, data sources, etc. etc. of the reports themselves to work with, this means that obtaining information on how each report works can be obtained right away or it can take a long time since, in the same way, there is no specific documentation that contains all this information, on the other hand Being a large number of reports, the option of taking a report as the basis for the project test was chosen, that is, a specific report was taken to test the techniques that are to be applied to solve the problem.

It has been agreed that certain reports will be taken for this solution proposal, for which guidelines have been created so that the reports are adequate to work with and that in the same way the product leaders have been discussed so that they can specify said reports that exist. at the moment.

- 1. Reports should not be Operational (Transactional, day to day, report with current information)

  - Oriented to customer service (taxpayer)
  - For the daily delivery of the departments
  - Reports of the day
  - Monitoring of requests attention flows

- 2. Must be Analytical reports (Not transactional, historical, not current day)

  - Weekly, Monthly, Annual Reports
  - Information reports sent to the Federation (information transparency)
  - Compendium reports from various sources of information (tables)
  - Reports that do not include the current day in their filter.

Analytical reports are geared towards supporting senior management's strategic and planning functions. Operational reporting is geared toward supporting day-to-day organizational functions. In the same way, it has been agreed on the functioning of the data collection through the ETL and on the other hand the visualization of the reports.

- 1. Information processing (ETL / Cube)

  - It would be during non-service hours
  - Not all of them should be executed daily, they should be the least
  - It must be taken into account that most databases have a daily maintenance plan (backup, log cleaning, etc...)
  - There would be an administration of the scheduling of the execution, by the DBA
  - Exceptionally, they could be executed on demand, under the risk of affecting the operation.
  - You might have the possibility to generate the PDF report after the execution of the processing.

- 2. Visualization:

  - They will be on the GRP, as currently shown, same display technology.
  - You just have to change the data source and redo the layout.

Currently there are personnel with knowledge in the process of the report "Mayor auxiliar de la cuenta", there is knowledge about the use of the report at the egob platform level, code, and related tables, for the moment it is known what the Store procedure of this report, but we do not have full knowledge about the operation of this file. There is staff to support the analysis of the entire operation of the report "Mayor auxiliar de la cuenta", in the same way there is access to a copy of the SIGG database from the beginning of the year 2022, to the repository RF- RM where the report flow code is located and to the egob platform for generating the report. The development of the project has been limited until the end of April 2022, since it is under the development of a data engineering resident, so it is intended to have the proposal for the solution of the problem for said established dates.

## 1.5 Objectives

### 1.5.1 General Objectives

It is intended that a proposal be presented on the data architecture for the report ”Mayor auxiliar de la cuenta” once the analysis of this report has been completed, which will be used as a base example for future reports. (Pipeline proposal).

### 1.5.2 Specific Objectives

- Analysis document of the report “Mayor auxiliar de la cuenta”.
- ETL structure proposal.
- Data Warehouse/OLAP Cube proposal.

# 2 Theoretical framework

## 2.1 Microsoft SQL Server Managment Studio

SQL Server Management Studio (SSMS) is an integrated environment for managing any SQL infrastructure. Use SSMS to access, configure, manage, administer, and develop all components of SQL Server, Azure SQL Database , Azure SQL Managed Instance, SQL Server on Azure VM, and Azure Synapse Analytics. SSMS provides a single comprehensive utility that combines a broad group of graphical tools with many rich script editors to provide access to SQL Server for developers and database administrators of all skill levels. [1]

## 2.2 Data Warehouse

A data warehouse is a type of data management system that is designed to enable and support business intelligence (BI) activities, especially analytics. Data warehouses are solely intended to perform queries and analysis and often contain large amounts of historical data. The data within a data warehouse is usually derived from a wide range of sources such as application log files and transaction applications. [2]

A data warehouse centralizes and consolidates large amounts of data from multiple sources. Its analytical capabilities allow organizations to derive valuable business insights from their data to improve decision-making. Over time, it builds a historical record that can be invaluable to data scientists and business analysts. Because of these capabilities, a data warehouse can be considered an organization's "single source of truth."

## 2.3 OLAP CUBE

Online analytical processing (OLAP) cubes are a feature in Service Manager that use the existing data warehouse infrastructure to provide self-service business intelligence capabilities to end users.

An OLAP cube is a data structure that overcomes the limitations of relational databases by providing rapid analysis of data. Cubes can display and sum large amounts of data while also providing users with searchable access to any data points. This way, the data can be rolled up, sliced, and diced as needed to handle the widest variety of questions that are relevant to a user's area of interest. [3]

## 2.4 Store Procedure

A stored procedure is a set of Structured Query Language (SQL) statements with an assigned name, which are stored in a relational database management system (RDBMS) as a group, so it can be reused and shared by multiple programs.

Stored procedures can access or modify data in a database, but it is not tied to a specific database or object, which offers a number of advantages. [4]

## 2.5 ETL

ETL, which stands for extract, transform and load, is a data integration process that combines data from multiple data sources into a single, consistent data store that is loaded into a data warehouse or other target system. [5]

## 2.6 Pipeline

In computing, pipeline refers to the logical queue that is filled with all the instructions for the computer processor to process in parallel. It is the process of storing and queuing tasks and instructions that are executed simultaneously by the processor in an organized way. [6]

## 2.7 SIGG

SISTEMA INTEGRAL DE GESTIÓN GUBERNAMENTAL - SIGG. Government reports of various types of nature, such as accounting, human resources, etc., are generated and registered on this platform.

## 2.8 EGOB

It is the comprehensive technological platform for the digital transformation of the Government information.

# 3 Methodology and development

During the month of January there was a complete introduction to the tools that were going to be used during this project which were SQL, Excel, SSIS. As a first approach before starting the analysis part, a pipeline proposal was generated that would follow the information flow process on the other hand, just as a proof of concept was created in a more general way of which would be the parts of this ETL process, below is the image of the pipeline that follows the entire flow of information and the parts in a general way.

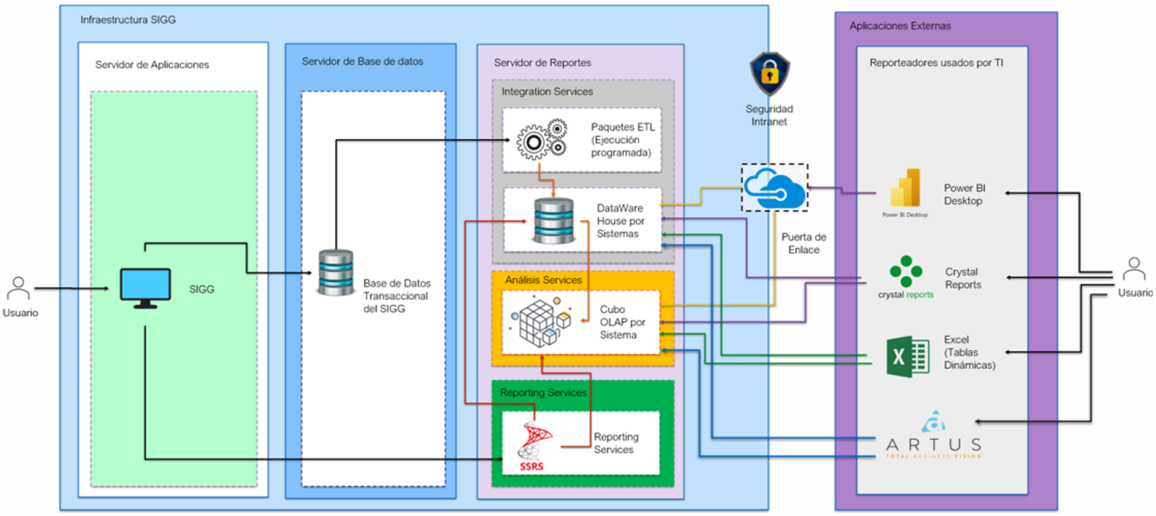


Figure 3.1: Pipeline

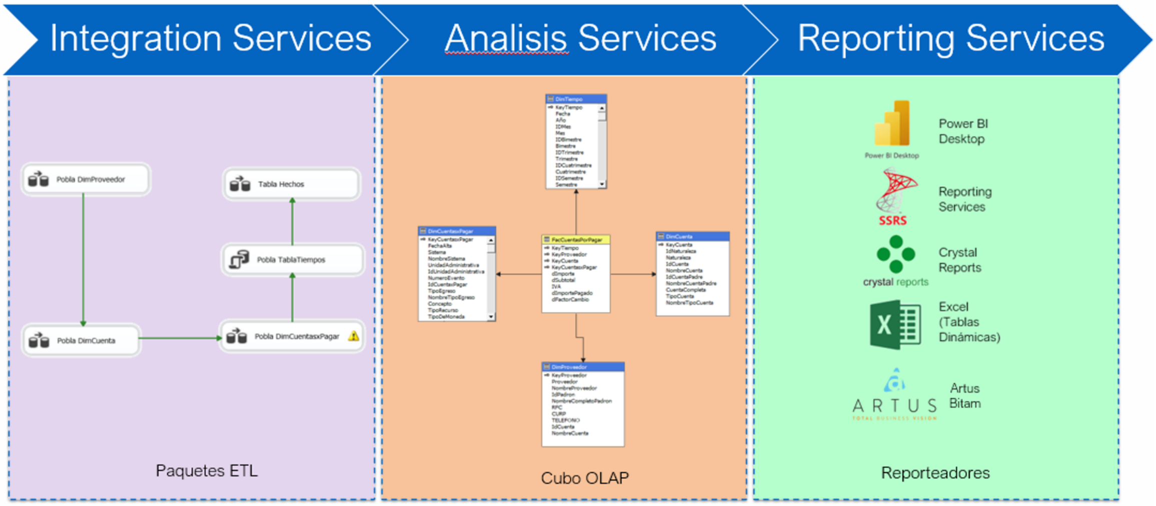


Figure 3.2: General Process

At the beginning of the month of February 2022, the development of the project aimed at government reports began under the supervision of Octavio Vallejo and Didier Moreno, a brief introduction to the problems was given and in the same way was shown all the resources that were available up to that moment, which consisted of a previous proposal with the same objective of reducing time in generating reports, therefore, once the problem was explained, I proceeded with the creation of some guidelines and restrictions to distinguish the types of reports to which this project was directed. The realization of this document lasted about 2 weeks since there was no solid knowledge on the part of me about the reports that are handled, for which the support of Octavio Vallejo was had for the realization of these points which have been mentioned in the scope and limitations section, it was concluded that the reports to which this project is directed are classified as analytical reports which are oriented towards the support of strategic functions and planning of senior management, that is, reports that handle historical data.

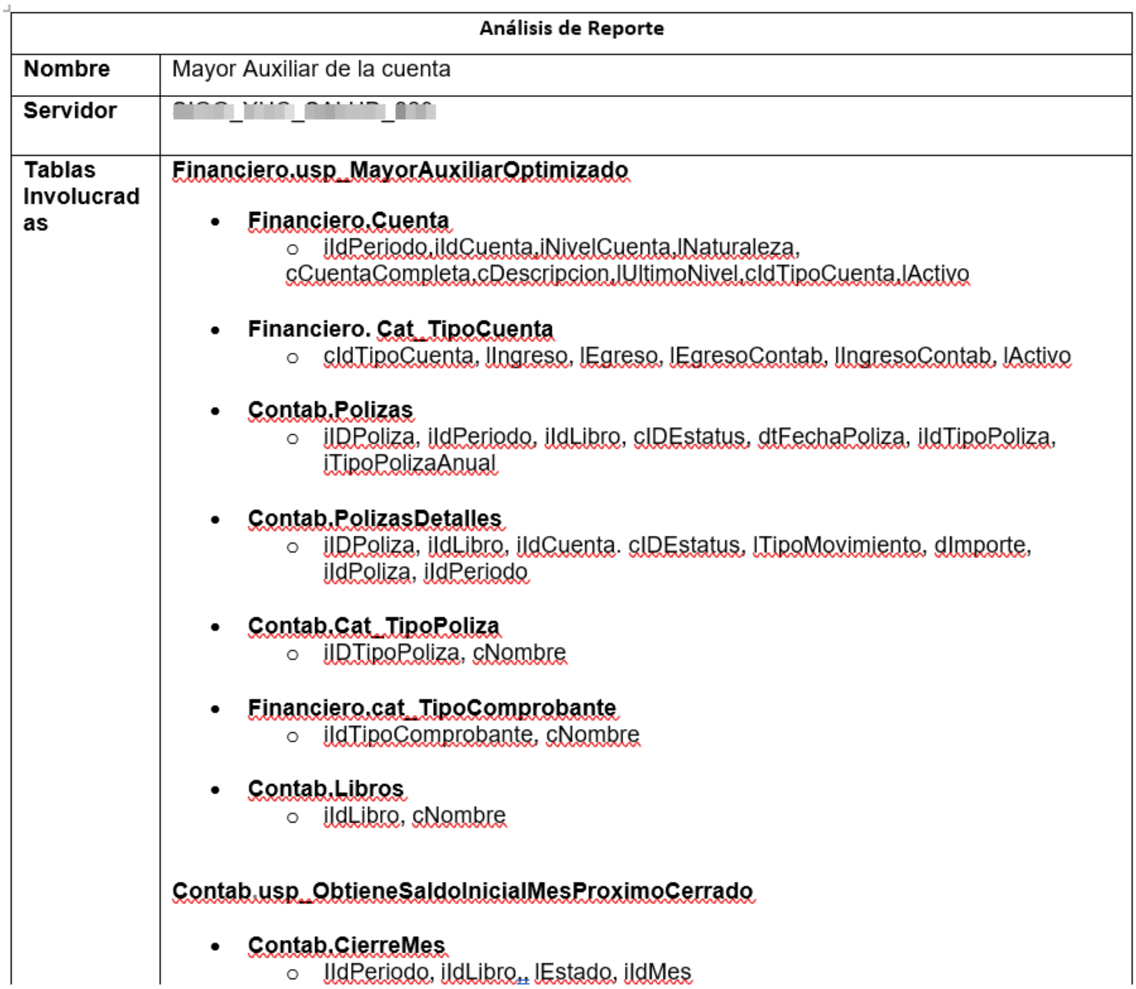
Once we had the guidelines and restrictions for the reports, we consulted with the product manager Ricardo Chi to support us with a list of report names that would fall within the previously established guidelines and restrictions, because this process would take between one and two weeks, he provided us with the name of a report which falls within the range of established reports of the project so that we could use it as a base example, that is, the report ”Mayor auxiliar de la cuenta”, it is the report that will be used for the test of concept for the solution of the established problematic. An objective planning scheme was carried out to be clear about the objectives to be carried out, taking as a sample the Mayor auxiliar de la cuenta. Three sections were established for the entire process that would be followed on the project.

- Data collection.
- Information processing
- Publication of results

## 3.1 Data collection

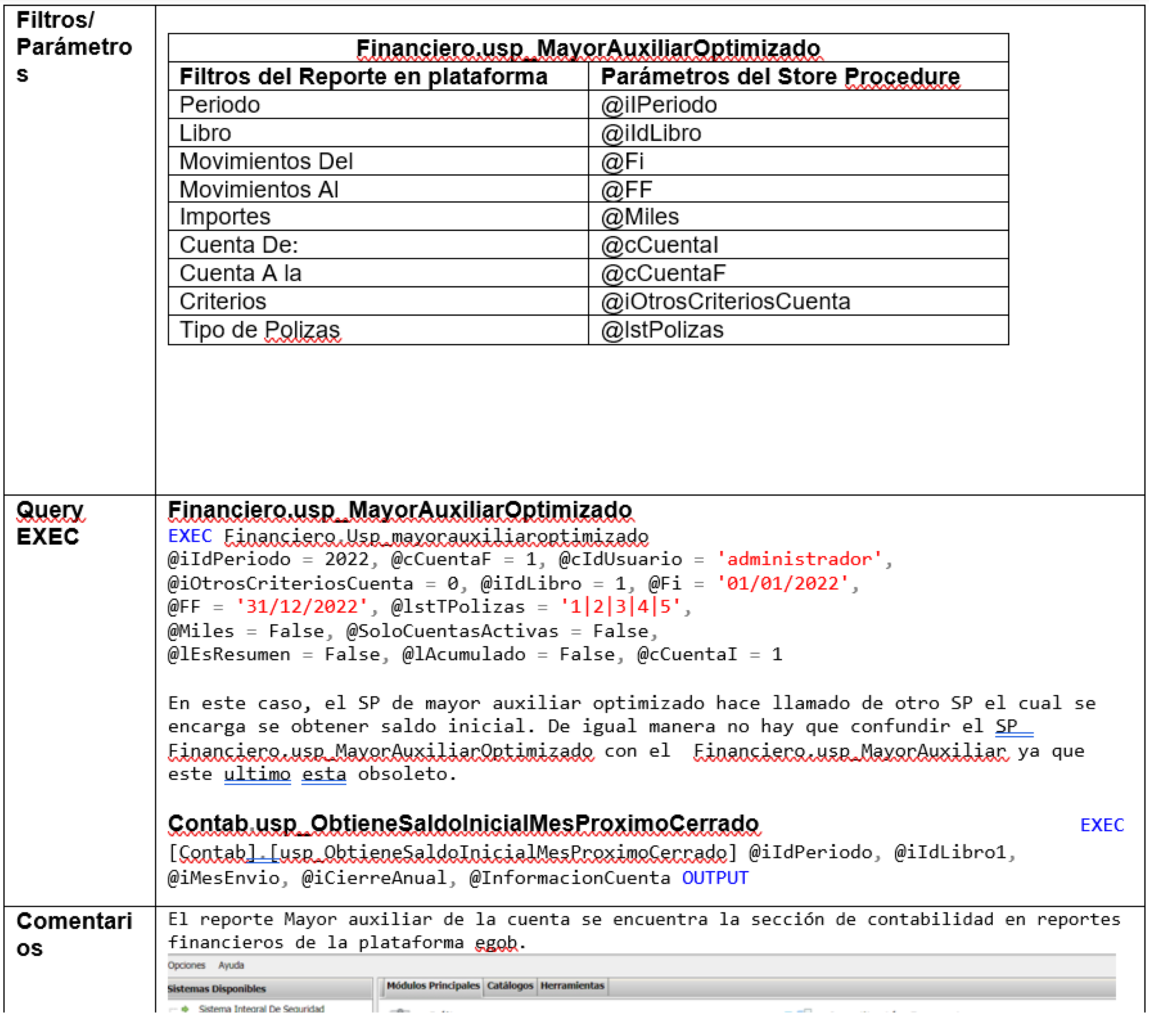
During the data collection, meetings were made for the full understanding of the report ”Mayor auxiliar de la cuenta” and in the same way the necessary permissions were requested to access the database, the repository and the report generation platform, several meetings were made with the product leader Manuel Manrique for the explanation of the entire operation of the report, from its example of use in the egob platform to the operation at the database level. On the other hand, there was an approach with the technical leader Elias Perez for the understanding of the flow of information at the code level, the Stored Procedure was found that calculates and filters information for the generation of the report and in the same way the crystal report schema used to display the same report was found. Based on this information obtained, a document was started on the analysis of the report ”Mayor auxiliar de la cuenta” in which all the detailed information about this report, its location, at the code level, database, related tables, etc. This process of meetings and explanations about the operation of the report lasted about two weeks since for work reasons the time dedicated by the leader and product manager was compromised and had to be delayed more than once.

On the other hand, the analysis of the operation of the Store Procedure of the report, ”Financiero.Usp-mayorauxiliaroptimizado”, was carried out, this process was extended more than planned since there is no personnel who is familiar with the use of this file, in this case it was analyzed and as a result the name of the tables used for the calculation and filtering of the report was obtained, in the same way the support of the software engineer Ana Tec was obtained for the generation of the query to make the call of the Store Procedure at the base level of data. At that time there was a copy of the SIGG database ”SIGG-YUC-SALUD-820”, it was requested that it be updated to the date of 2022 since a more recent version of the store procedure that would be analyzed was located here. This is an example of a little part of the document:



| **Análisis de Reporte** | |
|---|---|
| **Nombre** | Mayor Auxiliar de la cuenta |
| **Servidor** | [redacted] |
| **Tablas Involucradas** | **Financiero.usp_MayorAuxiliarOptimizado**<br>• **Financiero.Cuenta**<br>○ iIdPeriodo,iIdCuenta,iNivelCuenta,lNaturaleza, cCuentaCompleta,cDescripcion,lUltimoNivel,cIdTipoCuenta,lActivo<br>• **Financiero. Cat_TipoCuenta**<br>○ cIdTipoCuenta, lIngreso, lEgreso, lEgresoContab, lIngresoContab, lActivo<br>• **Contab.Polizas**<br>○ iIDPoliza, iIdPeriodo, iIdLibro, cIDEstatus, dtFechaPoliza, iIdTipoPoliza, iTipoPolizaAnual<br>• **Contab.PolizasDetalles**<br>○ iIDPoliza, iIdLibro, iIdCuenta. cIDEstatus, lTipoMovimiento, dImporte, iIdPoliza, iIdPeriodo<br>• **Contab.Cat_TipoPoliza**<br>○ iIDTipoPoliza, cNombre<br>• **Financiero.cat_TipoComprobante**<br>○ iIdTipoComprobante, cNombre<br>• **Contab.Libros**<br>○ iIdLibro, cNombre<br>**Contab.usp_ObtieneSaldoInicialMesProximoCerrado**<br>• **Contab.CierreMes**<br>○ lIdPeriodo, iIdLibro,, lEstado, iIdMes |

Figure 3.3: Analytical report P1



**Filtros/ Parámetros**

| Financiero.usp_MayorAuxiliarOptimizado | |
|---|---|
| **Filtros del Reporte en plataforma** | **Parámetros del Store Procedure** |
| Periodo | @iIPeriodo |
| Libro | @iIdLibro |
| Movimientos Del | @Fi |
| Movimientos Al | @FF |
| Importes | @Miles |
| Cuenta De: | @cCuentaI |
| Cuenta A la | @cCuentaF |
| Criterios | @iOtrosCriteriosCuenta |
| Tipo de Polizas | @lstPolizas |

**Query EXEC**

**Financiero.usp_MayorAuxiliarOptimizado**

```
EXEC Financiero.Usp_mayorauxiliaroptimizado
@iIdPeriodo = 2022, @cCuentaF = 1, @cIdUsuario = 'administrador',
@iOtrosCriteriosCuenta = 0, @iIdLibro = 1, @Fi = '01/01/2022',
@FF = '31/12/2022', @lstTPolizas = '1|2|3|4|5',
@Miles = False, @SoloCuentasActivas = False,
@lEsResumen = False, @lAcumulado = False, @cCuentaI = 1
```

En este caso, el SP de mayor auxiliar optimizado hace llamado de otro SP el cual se encarga se obtener saldo inicial. De igual manera no hay que confundir el SP Financiero.usp_MayorAuxiliarOptimizado con el Financiero.usp_MayorAuxiliar ya que este ultimo esta obsoleto.

**Contab.usp_ObtieneSaldoInicialMesProximoCerrado** EXEC

```
[Contab].[usp_ObtieneSaldoInicialMesProximoCerrado] @iIdPeriodo, @iIdLibro1,
@iMesEnvio, @iCierreAnual, @InformacionCuenta OUTPUT
```

**Comentarios**

El reporte Mayor auxiliar de la cuenta se encuentra la sección de contabilidad en reportes financieros de la plataforma egob.

Figure 3.4: Analytical report P2

## 3.2 Information processing

As a first approach, the cube concept began to be molded, I investigated on several websites how to work from scratch an OLAP cube, after investigating I made the first sketch of a cube dedicated to the report with which we are working, this first version of cube it was a snowflake model, in this case, it was rejected since it was not intended to be this way, but rather it had to have denormalized data, this version was discarded and the development of a star version began.

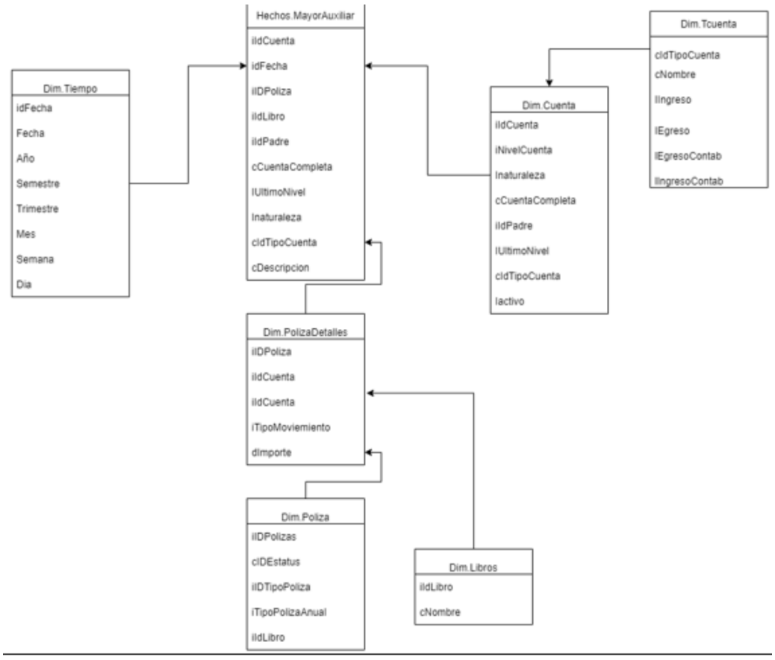


Figure 3.5: Snowflake Model

In the same way, work began on the first concept of what the ETL flow would be, so it was thought about an on-demand service that, after being analyzed, was discarded since in this case the waiting time for generating reports could being greater than what we already have up to now, basically we were replacing the "SP" that we already had with an ETL.

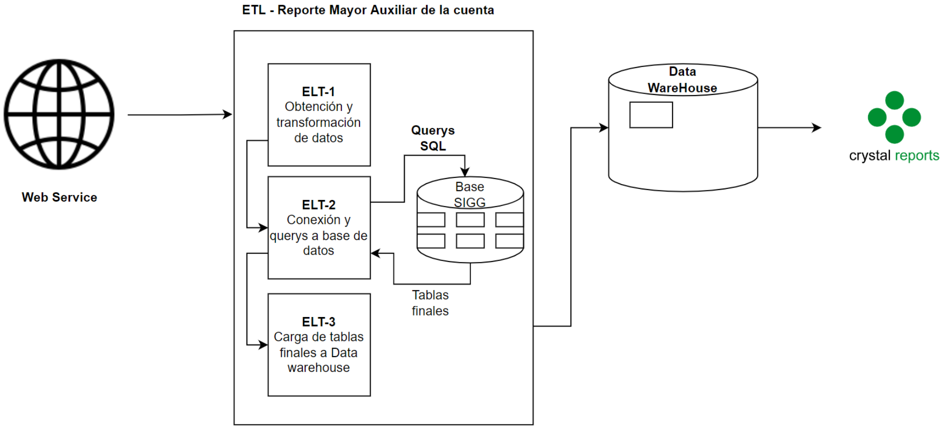

ETL - Reporte Mayor Auxiliar de la cuenta
Web Service
ELT-1
Obtención y transformación de datos
ELT-2
Conexión y querys a base de datos
ELT-3
Carga de tablas finales a Data warehouse
Querys SQL
Base SIGG
Tablas finales
Data WareHouse
crystal reports

Figure 3.6: On-demand Service

In this version it was specified that the ETL was going to collect the data sent by the client, once this data was obtained, the ELT would transform the data and through queries it would connect to the transactional database to make the filters and calculations. necessary, finally the result would be temporary tables that would be the same as those of the data warehouse, so the ETL would only have to make the connection between tables and columns to upload them to the data warehouse and with that it would be sent to the crystal reports reporter, the problem of this version is that the service was on demand, so the problem of the waiting time would not be resolved, but it could increase the waiting time.

Once the necessary information on the operation and location of tables related to the report was collected, the approach of the cube scheme for the report began with another approach, it was decided to take the star model for the generation of the OLAP Cube, in this case it was passed by several weeks polishing the model, several meetings were held with the product leader to explain and answer questions about the needs of the report and its operation. The first version of the cube with a star model was reached.

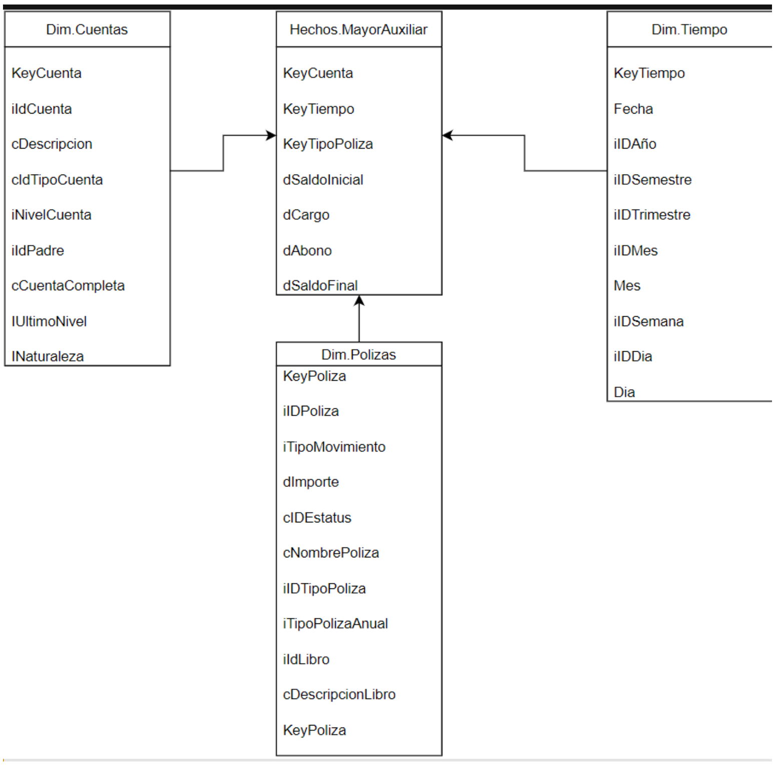
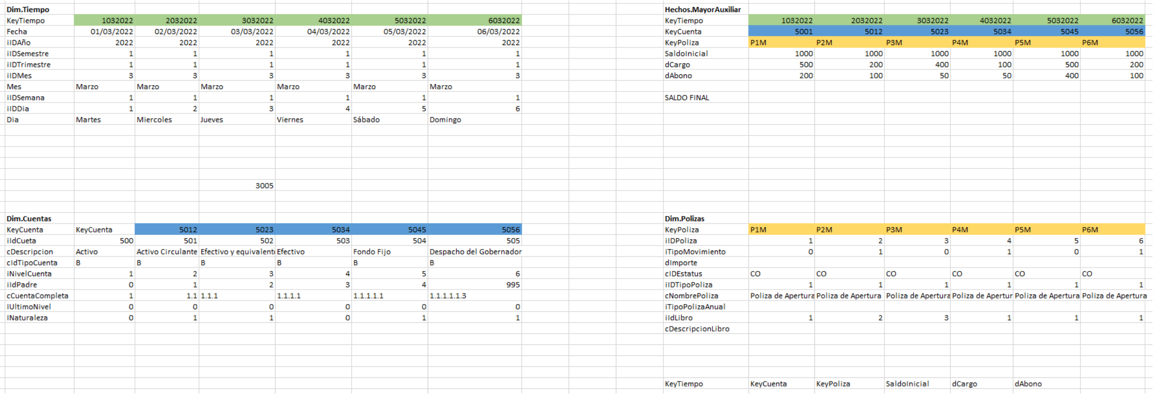


Figure 3.7: Star Model V1

For this version, it was tested for the first time with random data, that is, it was a controlled test to simulate the operation of the cube, for this process it was concluded that it was better to test with real data obtained from the database, data that is used to generate the report, all this simulation process was carried out on Excel.



| **Dim.Tiempo** | | | | | | |
|---|---|---|---|---|---|---|
| KeyTiempo | 1032022 | 2032022 | 3032022 | 4032022 | 5032022 | 6032022 |
| Fecha | 01/03/2022 | 02/03/2022 | 03/03/2022 | 04/03/2022 | 05/03/2022 | 06/03/2022 |
| iIDAño | 2022 | 2022 | 2022 | 2022 | 2022 | 2022 |
| iIDSemestre | 1 | 1 | 1 | 1 | 1 | 1 |
| iIDTrimestre | 1 | 1 | 1 | 1 | 1 | 1 |
| iIDMes | 3 | 3 | 3 | 3 | 3 | 3 |
| Mes | Marzo | Marzo | Marzo | Marzo | Marzo | Marzo |
| iIDSemana | 1 | 1 | 1 | 1 | 1 | 1 |
| iIDDia | 1 | 2 | 3 | 4 | 5 | 6 |
| Dia | Martes | Miercoles | Jueves | Viernes | Sábado | Domingo |
| | | | 3005 | | | |

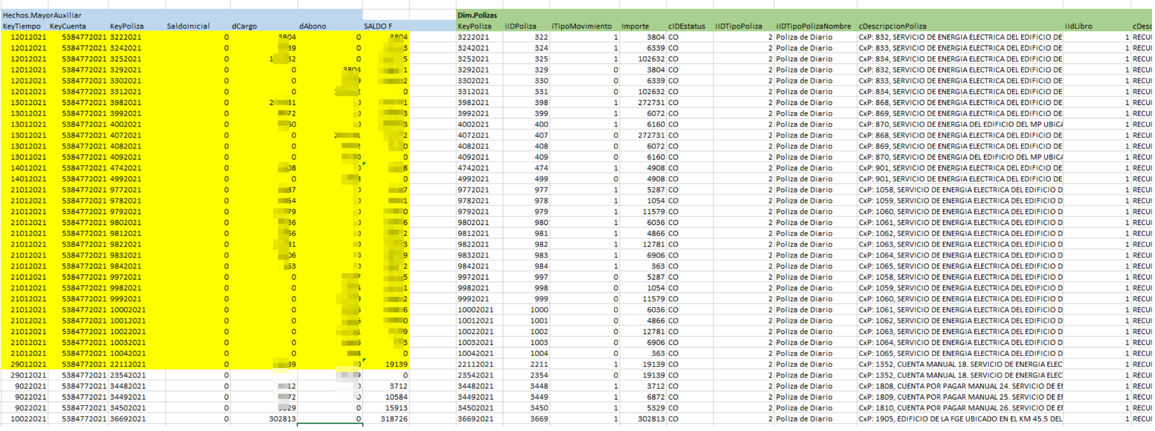
| **Dim.Cuentas** | | | | | | |
|---|---|---|---|---|---|---|
| KeyCuenta | KeyCuenta | 5012 | 5023 | 5034 | 5045 | 5056 |
| iIdCueta | 500 | 501 | 502 | 503 | 504 | 505 |
| cDescripcion | Activo | Activo Circulante | Efectivo y equivalent | Efectivo | Fondo Fijo | Despacho del Gobernador |
| cIdTipoCuenta | B | B | B | B | B | B |
| iNivelCuenta | 1 | 2 | 3 | 4 | 5 | 6 |
| iIdPadre | 0 | 1 | 2 | 3 | 4 | 995 |
| cCuentaCompleta | 1 | 1.1 | 1.1.1 | 1.1.1.1 | 1.1.1.1.1 | 1.1.1.1.1.3 |
| lUltimoNivel | 0 | 0 | 0 | 0 | 0 | 0 |
| lNaturaleza | 0 | 1 | 1 | 0 | 1 | 1 |

| **Hechos.MayorAuxiliar** | | | | | | |
|---|---|---|---|---|---|---|
| KeyTiempo | 1032022 | 2032022 | 3032022 | 4032022 | 5032022 | 6032022 |
| KeyCuenta | 5001 | 5012 | 5023 | 5034 | 5045 | 5056 |
| KeyPoliza | P1M | P2M | P3M | P4M | P5M | P6M |
| SaldoInicial | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| dCargo | 500 | 200 | 400 | 100 | 500 | 200 |
| dAbono | 200 | 100 | 50 | 50 | 400 | 100 |
| | | | | | | |
| SALDO FINAL | | | | | | |

| **Dim.Polizas** | | | | | | |
|---|---|---|---|---|---|---|
| KeyPoliza | P1M | P2M | P3M | P4M | P5M | P6M |
| iIDPoliza | 1 | 2 | 3 | 4 | 5 | 6 |
| iTipoMovimiento | 0 | 1 | 0 | 1 | 0 | 1 |
| dImporte | | | | | | |
| cIDEstatus | CO | CO | CO | CO | CO | CO |
| iIDTipoPoliza | 1 | 1 | 1 | 1 | 1 | 1 |
| cNombrePoliza | Poliza de Apertura | Poliza de Apertura | Poliza de Apertura | Poliza de Apertur | Poliza de Apertura | Poliza de Apertura |
| iTipoPolizaAnual | | | | | | |
| iIdLibro | 1 | 2 | 3 | 1 | 1 | 1 |
| cDescripcionLibro | | | | | | |
| | | | | | | |
| KeyTiempo | KeyCuenta | KeyPoliza | SaldoInicial | dCargo | dAbono | |

Figure 3.8: Firs Test

For the second test, columns were added that were not taken into account at the beginning, in the same way, the account "8.2.5.3000.3100.3110.3111.10.1.2.11.2.1.15317" was used as an example for the operation of the cube, the time of a full year to obtain data and the result of the query was compared at the database level and at the egob platform level, during this process, all the information of this account was collected and a test was made in Excel simulating how it would function the cube. I can not show all the data for reasons of confidentiality.



Hechos.MayorAuxiliar

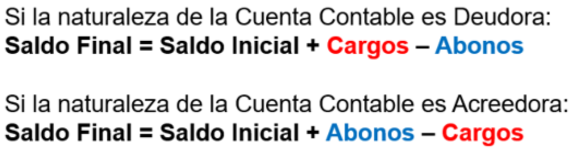
| KeyTiempo | KeyCuenta | KeyPoliza | SaldoInicial | dCargo | dAbono | SALDO F |
|---|---|---|---|---|---|---|
| 12012021 | 5384772021 | 3222021 | 0 | 3804 | 0 | 3804 |
| 12012021 | 5384772021 | 3242021 | 0 | [redacted] | 0 | [redacted] |
| 12012021 | 5384772021 | 3252021 | 0 | [redacted] | 0 | [redacted] |
| 12012021 | 5384772021 | 3292021 | 0 | 0 | 3804 | [redacted] |
| 12012021 | 5384772021 | 3302021 | 0 | 0 | [redacted] | [redacted] |
| 12012021 | 5384772021 | 3312021 | 0 | 0 | [redacted] | 0 |
| 13012021 | 5384772021 | 3982021 | 0 | [redacted] | [redacted] | [redacted] |
| 13012021 | 5384772021 | 3992021 | 0 | [redacted] | [redacted] | [redacted] |
| 13012021 | 5384772021 | 4002021 | 0 | [redacted] | [redacted] | [redacted] |
| 13012021 | 5384772021 | 4072021 | 0 | 0 | [redacted] | [redacted] |
| 13012021 | 5384772021 | 4082021 | 0 | 0 | [redacted] | [redacted] |
| 13012021 | 5384772021 | 4092021 | 0 | 0 | [redacted] | 0 |
| 14012021 | 5384772021 | 4742021 | 0 | [redacted] | [redacted] | [redacted] |
| 14012021 | 5384772021 | 4992021 | 0 | 0 | [redacted] | 0 |
| 21012021 | 5384772021 | 9772021 | 0 | [redacted] | [redacted] | [redacted] |
| 21012021 | 5384772021 | 9782021 | 0 | [redacted] | [redacted] | [redacted] |
| 21012021 | 5384772021 | 9792021 | 0 | [redacted] | [redacted] | [redacted] |
| 21012021 | 5384772021 | 9802021 | 0 | [redacted] | [redacted] | [redacted] |
| 21012021 | 5384772021 | 9812021 | 0 | [redacted] | [redacted] | [redacted] |
| 21012021 | 5384772021 | 9822021 | 0 | [redacted] | [redacted] | [redacted] |
| 21012021 | 5384772021 | 9832021 | 0 | [redacted] | [redacted] | [redacted] |
| 21012021 | 5384772021 | 9842021 | 0 | [redacted] | [redacted] | [redacted] |
| 21012021 | 5384772021 | 9972021 | 0 | 0 | [redacted] | [redacted] |
| 21012021 | 5384772021 | 9982021 | 0 | 0 | [redacted] | [redacted] |
| 21012021 | 5384772021 | 9992021 | 0 | 0 | [redacted] | [redacted] |
| 21012021 | 5384772021 | 10002021 | 0 | 0 | [redacted] | [redacted] |
| 21012021 | 5384772021 | 10012021 | 0 | 0 | [redacted] | [redacted] |
| 21012021 | 5384772021 | 10022021 | 0 | 0 | [redacted] | [redacted] |
| 21012021 | 5384772021 | 10032021 | 0 | 0 | [redacted] | [redacted] |
| 21012021 | 5384772021 | 10042021 | 0 | 0 | [redacted] | 0 |
| 29012021 | 5384772021 | 22112021 | 0 | [redacted] | [redacted] | 19139 |
| 29012021 | 5384772021 | 23542021 | 0 | [redacted] | [redacted] | 0 |
| 9022021 | 5384772021 | 34482021 | 0 | [redacted] | [redacted] | 3712 |
| 9022021 | 5384772021 | 34492021 | 0 | [redacted] | [redacted] | 10584 |
| 9022021 | 5384772021 | 34502021 | 0 | [redacted] | 0 | 15913 |
| 10022021 | 5384772021 | 36692021 | 0 | 302813 | 0 | 318726 |

Dim.Polizas

| KeyPoliza | iIDPoliza | iTipoMovimiento | Importe | cIDEstatus | iIDTipoPoliza | iIDTipoPolizaNombre | cDescripcionPoliza | iIdLibro | cDesc |
|---|---|---|---|---|---|---|---|---|---|
| 3222021 | 322 | 1 | 3804 | CO | 2 | Poliza de Diario | CxP: 832, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO DE | 1 | RECU |
| 3242021 | 324 | 1 | 6339 | CO | 2 | Poliza de Diario | CxP: 833, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO DE | 1 | RECU |
| 3252021 | 325 | 1 | 102632 | CO | 2 | Poliza de Diario | CxP: 834, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO DE | 1 | RECU |
| 3292021 | 329 | 0 | 3804 | CO | 2 | Poliza de Diario | CxP: 832, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO DE | 1 | RECU |
| 3302021 | 330 | 0 | 6339 | CO | 2 | Poliza de Diario | CxP: 833, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO DE | 1 | RECU |
| 3312021 | 331 | 0 | 102632 | CO | 2 | Poliza de Diario | CxP: 834, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO DE | 1 | RECU |
| 3982021 | 398 | 1 | 272731 | CO | 2 | Poliza de Diario | CxP: 868, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO DE | 1 | RECU |
| 3992021 | 399 | 1 | 6072 | CO | 2 | Poliza de Diario | CxP: 869, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO DE | 1 | RECU |
| 4002021 | 400 | 1 | 6160 | CO | 2 | Poliza de Diario | CxP: 870, SERVICIO DE ENERGIA DEL EDIFICIO DEL MP UBICA | 1 | RECU |
| 4072021 | 407 | 0 | 272731 | CO | 2 | Poliza de Diario | CxP: 868, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO DE | 1 | RECU |
| 4082021 | 408 | 0 | 6072 | CO | 2 | Poliza de Diario | CxP: 869, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO DE | 1 | RECU |
| 4092021 | 409 | 0 | 6160 | CO | 2 | Poliza de Diario | CxP: 870, SERVICIO DE ENERGIA DEL EDIFICIO DEL MP UBICA | 1 | RECU |
| 4742021 | 474 | 1 | 4908 | CO | 2 | Poliza de Diario | CxP: 901, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO DE | 1 | RECU |
| 4992021 | 499 | 0 | 4908 | CO | 2 | Poliza de Diario | CxP: 901, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO DE | 1 | RECU |
| 9772021 | 977 | 1 | 5287 | CO | 2 | Poliza de Diario | CxP: 1058, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 9782021 | 978 | 1 | 1054 | CO | 2 | Poliza de Diario | CxP: 1059, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 9792021 | 979 | 1 | 11579 | CO | 2 | Poliza de Diario | CxP: 1060, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 9802021 | 980 | 1 | 6036 | CO | 2 | Poliza de Diario | CxP: 1061, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 9812021 | 981 | 1 | 4866 | CO | 2 | Poliza de Diario | CxP: 1062, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 9822021 | 982 | 1 | 12781 | CO | 2 | Poliza de Diario | CxP: 1063, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 9832021 | 983 | 1 | 6906 | CO | 2 | Poliza de Diario | CxP: 1064, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 9842021 | 984 | 1 | 363 | CO | 2 | Poliza de Diario | CxP: 1065, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 9972021 | 997 | 0 | 5287 | CO | 2 | Poliza de Diario | CxP: 1058, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 9982021 | 998 | 0 | 1054 | CO | 2 | Poliza de Diario | CxP: 1059, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 9992021 | 999 | 0 | 11579 | CO | 2 | Poliza de Diario | CxP: 1060, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 10002021 | 1000 | 0 | 6036 | CO | 2 | Poliza de Diario | CxP: 1061, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 10012021 | 1001 | 0 | 4866 | CO | 2 | Poliza de Diario | CxP: 1062, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 10022021 | 1002 | 0 | 12781 | CO | 2 | Poliza de Diario | CxP: 1063, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 10032021 | 1003 | 0 | 6906 | CO | 2 | Poliza de Diario | CxP: 1064, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 10042021 | 1004 | 0 | 363 | CO | 2 | Poliza de Diario | CxP: 1065, SERVICIO DE ENERGIA ELECTRICA DEL EDIFICIO D | 1 | RECU |
| 22112021 | 2211 | 1 | 19139 | CO | 2 | Poliza de Diario | CxP: 1352, CUENTA MANUAL 18. SERVICIO DE ENERGIA ELEC | 1 | RECU |
| 23542021 | 2354 | 0 | 19139 | CO | 2 | Poliza de Diario | CxP: 1352, CUENTA MANUAL 18. SERVICIO DE ENERGIA ELEC | 1 | RECU |
| 34482021 | 3448 | 1 | 3712 | CO | 2 | Poliza de Diario | CxP: 1808, CUENTA POR PAGAR MANUAL 24. SERVICIO DE E | 1 | RECU |
| 34492021 | 3449 | 1 | 6872 | CO | 2 | Poliza de Diario | CxP: 1809, CUENTA POR PAGAR MANUAL 25. SERVICIO DE E | 1 | RECU |
| 34502021 | 3450 | 1 | 5329 | CO | 2 | Poliza de Diario | CxP: 1810, CUENTA POR PAGAR MANUAL 26. SERVICIO DE E | 1 | RECU |
| 36692021 | 3669 | 1 | 302813 | CO | 2 | Poliza de Diario | CxP: 1905, EDIFICIO DE LA FGE UBICADO EN EL KM 45.5 DEL | 1 | RECU |

Figure 3.9: Second Test

For this point, it was added to the test that was carried out on the operation of the cube and the calculation in the same cube of the "Saldo Final" column, to obtain this digit the following formula was used.



Si la naturaleza de la Cuenta Contable es Deudora:
**Saldo Final = Saldo Inicial + Cargos – Abonos**

Si la naturaleza de la Cuenta Contable es Acreedora:
**Saldo Final = Saldo Inicial + Abonos – Cargos**

Figure 3.10: Saldo Final Operation

The operation of the entire reporting process was being analyzed in more depth and as a result it was concluded that it is possible to calculate the final balance and that it be reflected in the cube, which is why the second proposal of the cube emerged.

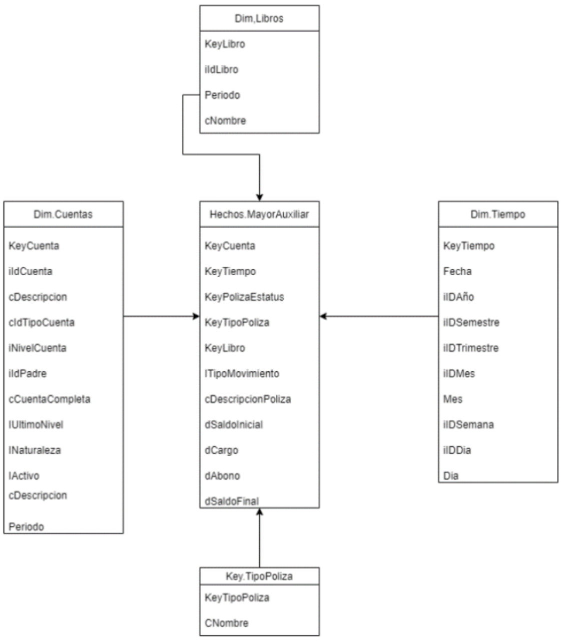


Figure 3.11: Star Model V2

In this case, the version of the cube was not approved since the calculation of these values (Saldo Final) was proposed through a Store Procedure. On this concept, it was proposed that a "SP" make certain calculations and information filtering and the client would simply take the necessary columns that would be obtained from a temporary table that would generate the same file.

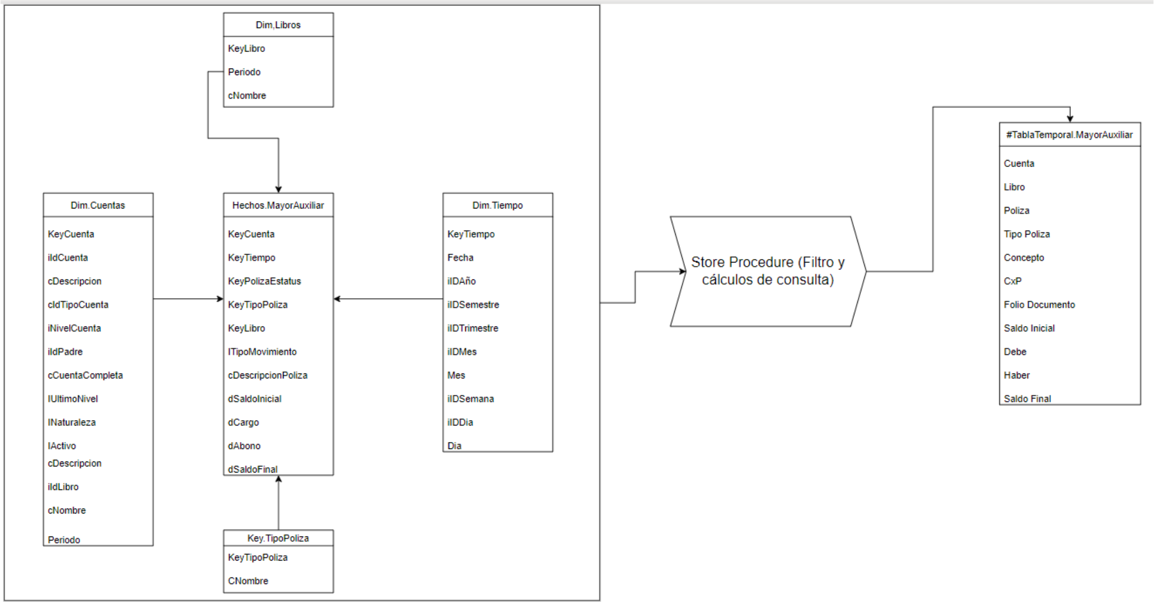


Figure 3.12: Star Model V2 with SP

At this point, this version was discarded since the goal is for the Cube to have all the data calculated, the client would simply get the necessary information with simple filter queries, no calculation would have to go through the client. With this conclusion, a third version of the cube was reached, which falls within those previously explained specifications. This version of the cube would have all the calculated information and the client would only have to filter, it would not use a Store procedure, it would simply be a process between ETL and Data warehouse.

This version was worked on for several months and after a lot of research, a more polished and different version of what was already available was reached.

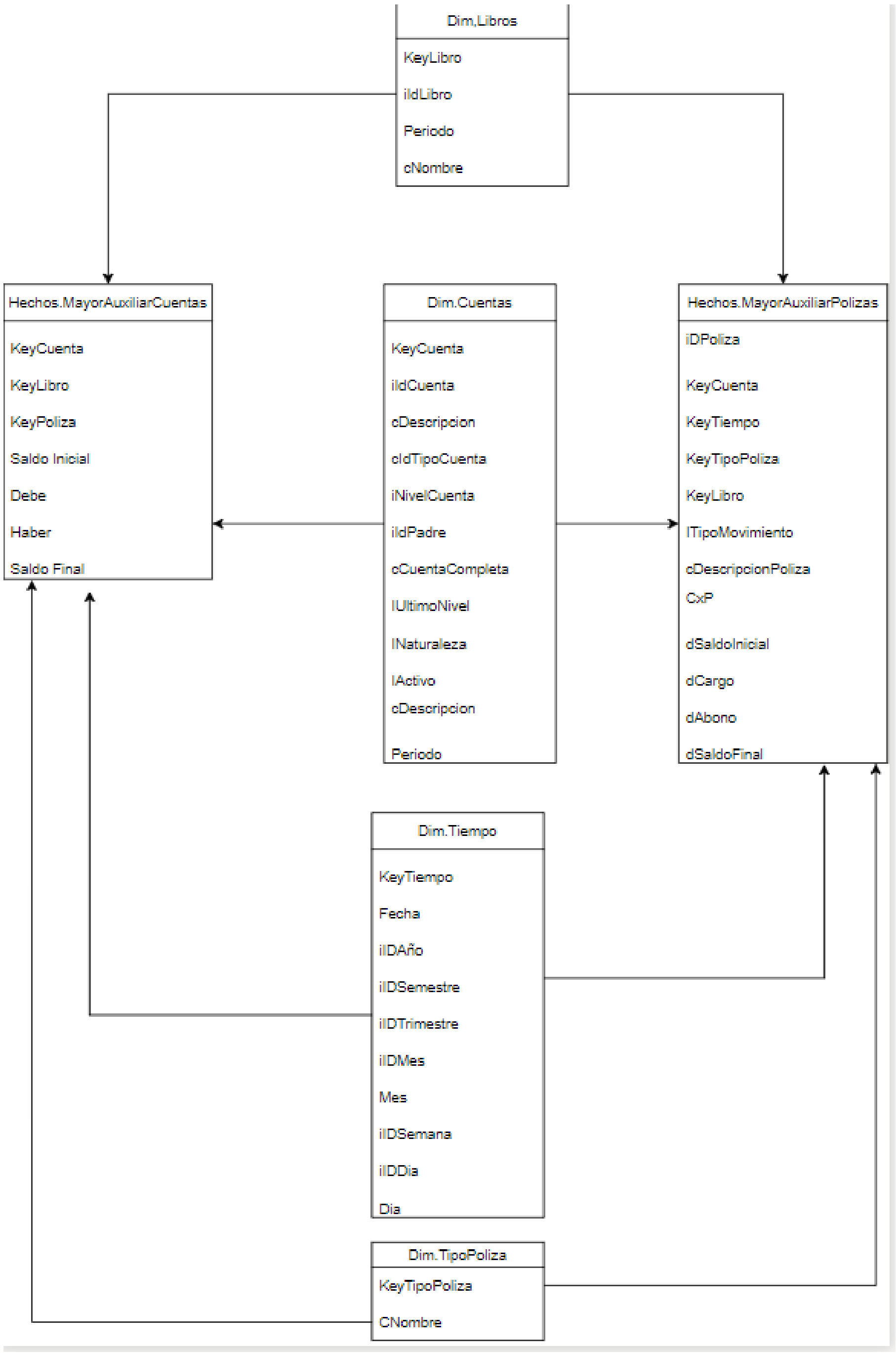
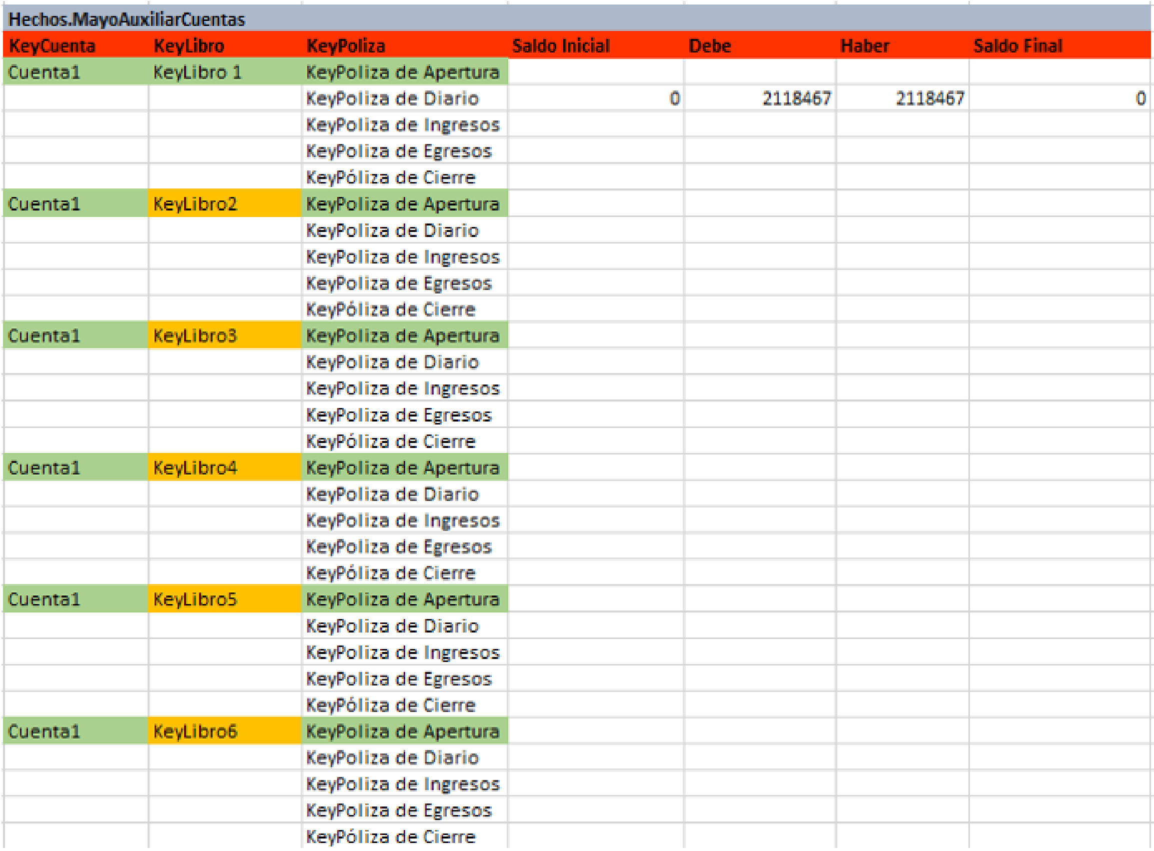


Figure 3.13: Constellation model

This model is known as a constellation scheme, in this case it has two fact tables which have different purposes, in the case of the fact table "MayorAuxiliarCuentas" there is a table in which all the accounts are held and they are combined with the books and types of policies.



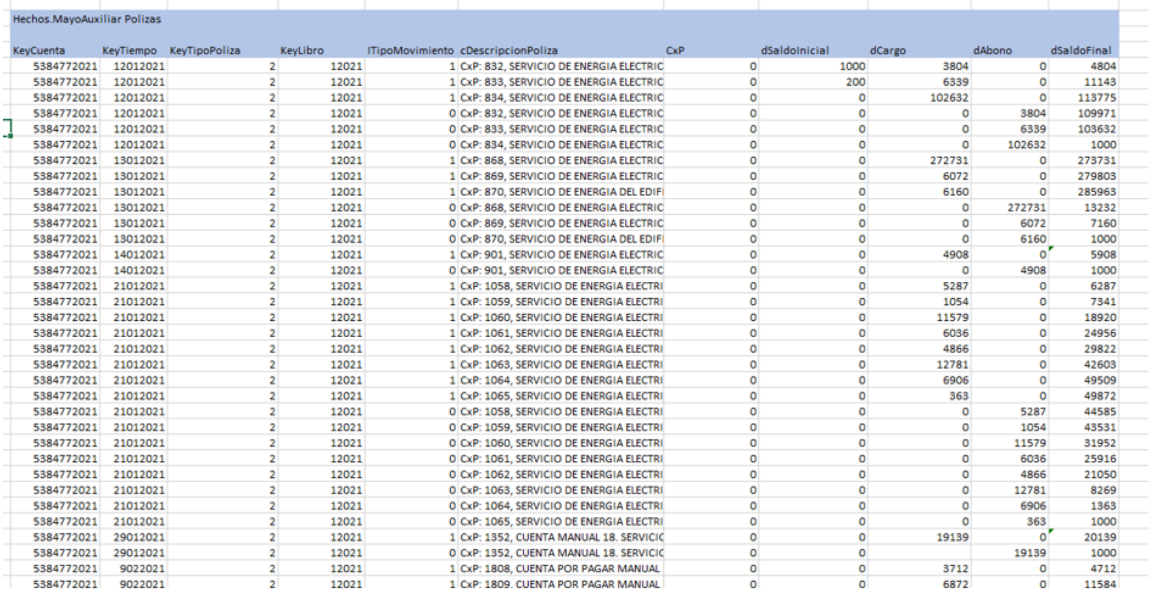
Hechos.MayoAuxiliarCuentas

| KeyCuenta | KeyLibro | KeyPoliza | Saldo Inicial | Debe | Haber | Saldo Final |
|---|---|---|---|---|---|---|
| Cuenta1 | KeyLibro 1 | KeyPoliza de Apertura | | | | |
| | | KeyPoliza de Diario | 0 | 2118467 | 2118467 | 0 |
| | | KeyPoliza de Ingresos | | | | |
| | | KeyPoliza de Egresos | | | | |
| | | KeyPóliza de Cierre | | | | |
| Cuenta1 | KeyLibro2 | KeyPoliza de Apertura | | | | |
| | | KeyPoliza de Diario | | | | |
| | | KeyPoliza de Ingresos | | | | |
| | | KeyPoliza de Egresos | | | | |
| | | KeyPóliza de Cierre | | | | |
| Cuenta1 | KeyLibro3 | KeyPoliza de Apertura | | | | |
| | | KeyPoliza de Diario | | | | |
| | | KeyPoliza de Ingresos | | | | |
| | | KeyPoliza de Egresos | | | | |
| | | KeyPóliza de Cierre | | | | |
| Cuenta1 | KeyLibro4 | KeyPoliza de Apertura | | | | |
| | | KeyPoliza de Diario | | | | |
| | | KeyPoliza de Ingresos | | | | |
| | | KeyPoliza de Egresos | | | | |
| | | KeyPóliza de Cierre | | | | |
| Cuenta1 | KeyLibro5 | KeyPoliza de Apertura | | | | |
| | | KeyPoliza de Diario | | | | |
| | | KeyPoliza de Ingresos | | | | |
| | | KeyPoliza de Egresos | | | | |
| | | KeyPóliza de Cierre | | | | |
| Cuenta1 | KeyLibro6 | KeyPoliza de Apertura | | | | |
| | | KeyPoliza de Diario | | | | |
| | | KeyPoliza de Ingresos | | | | |
| | | KeyPoliza de Egresos | | | | |
| | | KeyPóliza de Cierre | | | | |

Figure 3.14: Table test - Hechos.MayorAuxiliarCuenta

This is an example of how you would see the table, keep in mind that the filters used in the report are the account, the type of book and the type of policy, so the table is made up of a combination of these 3 main filters, in this case the values are already calculated through the ETL and simply the client would have to do a filter on this fact table, it is worth mentioning that this table would only be updated, it can be seen as a type of dynamic catalog that is updates according to the policies that are generated, on the other hand, being a combination of several filters, the size of this table is related to the number of accounts, books and types of policies that exist on the platform.

Regarding the "MayorAuxiliarPolizas" fact table, this table is aimed at individual policies, that is, there is a record of the policies one by one and they are also linked to dates.



Hechos.MayoAuxiliar Polizas

| KeyCuenta | KeyTiempo | KeyTipoPoliza | KeyLibro | ITipoMovimiento | cDescripcionPoliza | CxP | dSaldoInicial | dCargo | dAbono | dSaldoFinal |
|---|---|---|---|---|---|---|---|---|---|---|
| 5384772021 | 12012021 | 2 | 12021 | 1 | CxP: 832, SERVICIO DE ENERGIA ELECTRIC | 0 | 1000 | 3804 | 0 | 4804 |
| 5384772021 | 12012021 | 2 | 12021 | 1 | CxP: 833, SERVICIO DE ENERGIA ELECTRIC | 0 | 200 | 6339 | 0 | 11143 |
| 5384772021 | 12012021 | 2 | 12021 | 1 | CxP: 834, SERVICIO DE ENERGIA ELECTRIC | 0 | 0 | 102632 | 0 | 113775 |
| 5384772021 | 12012021 | 2 | 12021 | 0 | CxP: 832, SERVICIO DE ENERGIA ELECTRIC | 0 | 0 | 0 | 3804 | 109971 |
| 5384772021 | 12012021 | 2 | 12021 | 0 | CxP: 833, SERVICIO DE ENERGIA ELECTRIC | 0 | 0 | 0 | 6339 | 103632 |
| 5384772021 | 12012021 | 2 | 12021 | 0 | CxP: 834, SERVICIO DE ENERGIA ELECTRIC | 0 | 0 | 0 | 102632 | 1000 |
| 5384772021 | 13012021 | 2 | 12021 | 1 | CxP: 868, SERVICIO DE ENERGIA ELECTRIC | 0 | 0 | 272731 | 0 | 273731 |
| 5384772021 | 13012021 | 2 | 12021 | 1 | CxP: 869, SERVICIO DE ENERGIA ELECTRIC | 0 | 0 | 6072 | 0 | 279803 |
| 5384772021 | 13012021 | 2 | 12021 | 1 | CxP: 870, SERVICIO DE ENERGIA DEL EDIFI | 0 | 0 | 6160 | 0 | 285963 |
| 5384772021 | 13012021 | 2 | 12021 | 0 | CxP: 868, SERVICIO DE ENERGIA ELECTRIC | 0 | 0 | 0 | 272731 | 13232 |
| 5384772021 | 13012021 | 2 | 12021 | 0 | CxP: 869, SERVICIO DE ENERGIA ELECTRIC | 0 | 0 | 0 | 6072 | 7160 |
| 5384772021 | 13012021 | 2 | 12021 | 0 | CxP: 870, SERVICIO DE ENERGIA DEL EDIFI | 0 | 0 | 0 | 6160 | 1000 |
| 5384772021 | 14012021 | 2 | 12021 | 1 | CxP: 901, SERVICIO DE ENERGIA ELECTRIC | 0 | 0 | 4908 | 0 | 5908 |
| 5384772021 | 14012021 | 2 | 12021 | 0 | CxP: 901, SERVICIO DE ENERGIA ELECTRIC | 0 | 0 | 0 | 4908 | 1000 |
| 5384772021 | 21012021 | 2 | 12021 | 1 | CxP: 1058, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 5287 | 0 | 6287 |
| 5384772021 | 21012021 | 2 | 12021 | 1 | CxP: 1059, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 1054 | 0 | 7341 |
| 5384772021 | 21012021 | 2 | 12021 | 1 | CxP: 1060, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 11579 | 0 | 18920 |
| 5384772021 | 21012021 | 2 | 12021 | 1 | CxP: 1061, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 6036 | 0 | 24956 |
| 5384772021 | 21012021 | 2 | 12021 | 1 | CxP: 1062, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 4866 | 0 | 29822 |
| 5384772021 | 21012021 | 2 | 12021 | 1 | CxP: 1063, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 12781 | 0 | 42603 |
| 5384772021 | 21012021 | 2 | 12021 | 1 | CxP: 1064, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 6906 | 0 | 49509 |
| 5384772021 | 21012021 | 2 | 12021 | 1 | CxP: 1065, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 363 | 0 | 49872 |
| 5384772021 | 21012021 | 2 | 12021 | 0 | CxP: 1058, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 0 | 5287 | 44585 |
| 5384772021 | 21012021 | 2 | 12021 | 0 | CxP: 1059, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 0 | 1054 | 43531 |
| 5384772021 | 21012021 | 2 | 12021 | 0 | CxP: 1060, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 0 | 11579 | 31952 |
| 5384772021 | 21012021 | 2 | 12021 | 0 | CxP: 1061, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 0 | 6036 | 25916 |
| 5384772021 | 21012021 | 2 | 12021 | 0 | CxP: 1062, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 0 | 4866 | 21050 |
| 5384772021 | 21012021 | 2 | 12021 | 0 | CxP: 1063, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 0 | 12781 | 8269 |
| 5384772021 | 21012021 | 2 | 12021 | 0 | CxP: 1064, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 0 | 6906 | 1363 |
| 5384772021 | 21012021 | 2 | 12021 | 0 | CxP: 1065, SERVICIO DE ENERGIA ELECTRI | 0 | 0 | 0 | 363 | 1000 |
| 5384772021 | 29012021 | 2 | 12021 | 1 | CxP: 1352, CUENTA MANUAL 18. SERVICIC | 0 | 0 | 19139 | 0 | 20139 |
| 5384772021 | 29012021 | 2 | 12021 | 0 | CxP: 1352, CUENTA MANUAL 18. SERVICIC | 0 | 0 |  | 19139 | 1000 |
| 5384772021 | 9022021 | 2 | 12021 | 1 | CxP: 1808, CUENTA POR PAGAR MANUAL | 0 | 0 | 3712 | 0 | 4712 |
| 5384772021 | 9022021 | 2 | 12021 | 1 | CxP: 1809. CUENTA POR PAGAR MANUAL | 0 | 0 | 6872 | 0 | 11584 |

Figure 3.15: Table test - Hechos.MayorAuxiliarPolizas

The ETL is in charge of filling the cube with information obtained from the transactional database, in this case, in this process the ordering of data would be done and in the same way all the mathematical calculations every time the process of the ETL, this process is still to be defined since the consultation cycles of each client must be taken into account for the generation of the report.

In this case, as the report moves according to the policies, for the filling of information and calculation, a count of all the accounts, books and policies that exist in the database would be obtained as a first execution, taking into account from the period that was defined and in the same way the first tables to fill in would be those that carry information on the policies since they are the pivotal information of the report.

On the other hand, each fact table is linked to its dimensions, in this case, the account fact table is linked to an account dimension table in which the characteristics of the same are contemplated, in this way the father account with daughter account, in the same way if this account is active or not, if it is a debtor, etc. On the other hand, the size of the book and type of policy are taken into account for filtering through the ”Keys”.

# 4 Results

As a result of this project, the proposal of the cube was aimed at, it was decided to focus on the design of the cube and leave aside the structure of the ETL so that as a final result the schematic model is obtained, the testing of this model by means of real data in excel, in the same way the documentation of the analysis of the report "Mayor Auxiliary of the account" which will be useful for future employees who resume the project with this report since that way they do not have to start from scratch the analysis of this such a complex report.

# 5 Conclusions and recommendations

## 5.1 Conclusion

This project is a process that has to be invested a lot of time since the complexity is based on the knowledge that one has about the reporting system and the operation of each of the reports, in this case, this project is not finished , there is no staff that has in-depth knowledge of how the report works at the store procedure level, so analyzing it to understand how it worked, the main tables to obtain the data necessary for the operation of the report took a long time, in addition to Every doubt that arose, it was preferable to schedule a call with the product leader or manager, which could not be immediate since their availability could be affected.

On the other hand, as already mentioned, this project is not finished, the proposed cube is designed so that the client only makes filters, the ETL does all the calculations every time the data warehouse update process is executed, it is necessary to have Keep in mind that the model that is handled right now for this report is on demand, that is, it goes from an interface to a web service that calls a store procedure and returns temporary tables where the final data for the report is, in this case , this project seeks that in the same way it is not a process under command but rather that they are direct filtering queries so that the client only visualizes it in the report that he wants. For this project it is still necessary to generate the ETL which can be based on the project ”IS-Cubo-300421.zip” (an internal project of the company), this is a project where you can see an ETL for filling a data warehouse . In the same way, the final cube must be created on a dedicated database which will function as a data warehouse.

# Bibliography

[1] rothja. What is sql server management studio (ssms)? https://docs.microsoft.com/en-us/sql/ssms/sql-server-management-studio-ssms?view=sql-server-ver15, jan 2022.

[2] oracle. What is a data warehouse? https://www.oracle.com/database/what-is-a-data-warehouse/, jan 2020.

[3] v anesh. Overview of service manger olap cubes for advanced analytics. https://docs.microsoft.com/en-us/system-center/scsm/olap-cubes-overview?view=sc-sm-2022, aug 2020.

[4] Adam Hughes. stored procedure. https://www.techtarget.com/searchoracle/definition/stored-procedure, jan 2021.

[5] IBM Cloud Education. Etl (extract, transform, load). https://www.ibm.com/cloud/learn/etl#:~:text=ETL%2C%20which%20stands%20for%20extract,warehouse%20or%20other%20target%20system., apr 2020.

[6] Techopedia. Pipeline. https://www.techopedia.com/definition/5312/pipeline, jan 2017.