CASO PRÁCTICO FINAL

Maestría: Big Data y Ciencia de Datos

Módulo: Business Intelligence

Docente: Ing. Juan Carlos Benalcázar

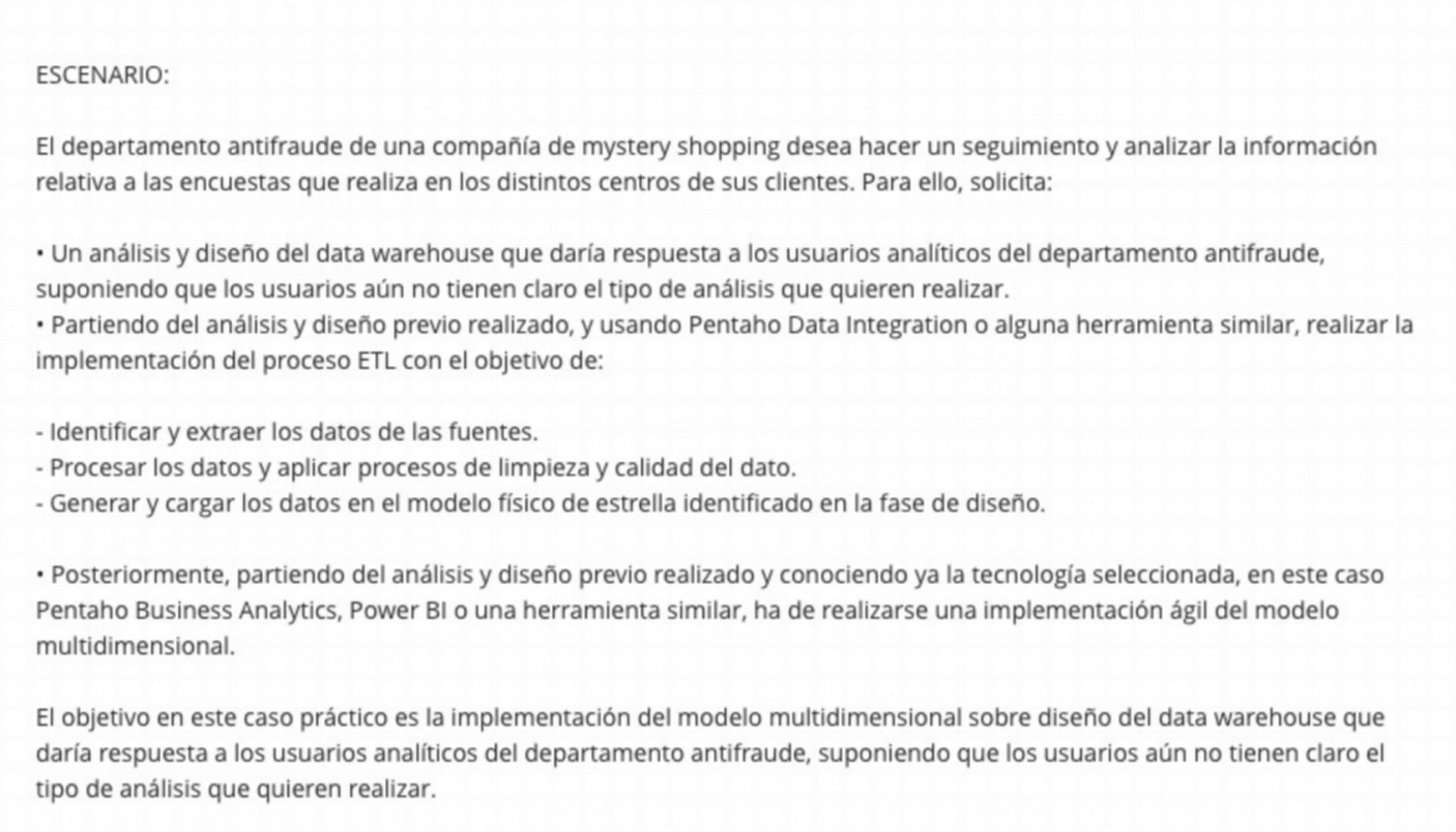
Integrantes: Cabezas Gutierrez Cristhian Fabián

Calle Revelo Gustavo John

Gaibor Rojas Sonia Marlene

Muñoz Montenegro Miguel Fernando

Fecha: Octubre 2022



Text, letter

Description automatically generated

1. **DESARROLLO**

El siguiente caso práctico se va a desarrollar en base al orden de los numerales planteados en el escenario:

1. **Análisis de fuentes de datos**

Se tiene como fuente de datos inicial al archivo IMF\_Mustery\_Shopping.csv que contiene ciertos datos de la Compañía Mystery Shopping ordenada por columnas. En términos generales este archivo contiene información de las evaluaciones realizadas en las diferentes provincias y localidades asignadas a un auditor con su respectivo resultado.

Haciendo un primer análisis de la fuente de datos referencial, se ha elaborado el siguiente diccionario de datos, que no es más que el conjunto de metadatos que enlistan las características lógicas y puntuales de los datos que, para el desarrollo de este caso práctico, se van a utilizar para las diferentes fases de la solución.

* 1. Descripción de detalle de cada campo

|  |  |
| --- | --- |
| Nombre del campo | Descripción |
| COD\_LOC | Guarda código de localidad |
| NOMBRE\_LOC | Guarda nombre de localidad |
| CP | Guarda código postal |
| POBLACION | Guarda nombre de población |
| OFICINA | Guarda código de oficina |
| PROVINCIA | Guarda nombre de provincia |
| COD\_PROY | Guarda código de proyecto |
| ID\_EVALUACION | Guarda id de evaluacion |
| Fecha de ejecucion | Guarda fecha de ejecución |
| COD\_AUDITOR | Guarda código de auditor |
| RESULTADO | Guarda resultado de evaluación |
| TITULO\_CUESTIONARIO | Guarda nombre del cuestionario |

*Tabla 1. Detalle de los campos*

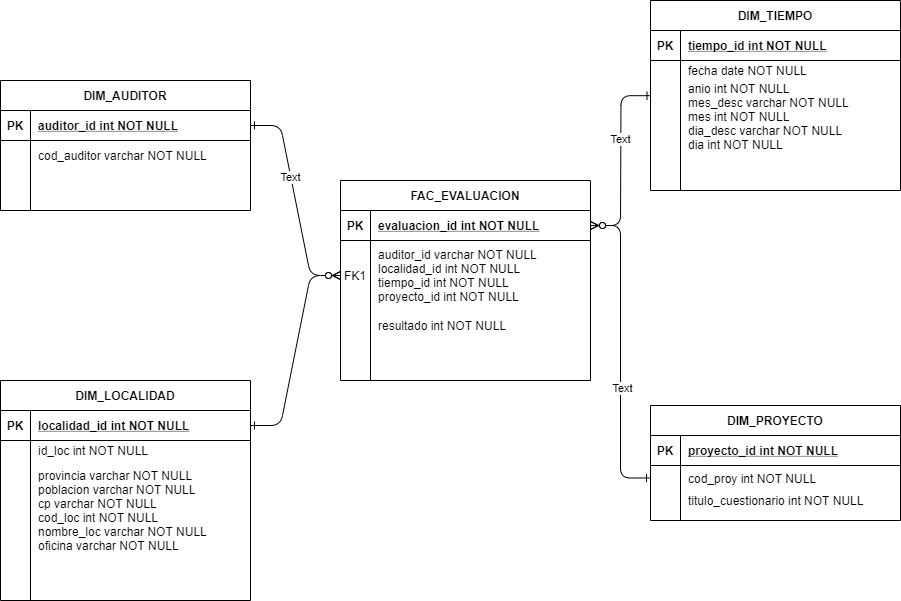
* 1. Tipo de campo, naturaleza

|  |  |
| --- | --- |
| Nombre del campo | Tipo de dato |
| COD\_LOC | varchar(15) |
| NOMBRE\_LOC | varchar(60) |
| CP | Number |
| POBLACION | varchar(50) |
| OFICINA | varchar(3) |
| PROVINCIA | varchar(25) |
| COD\_PROY | varchar(15) |
| ID\_EVALUACION | Number |
| Fecha de ejecucion | Date (DD/MM/YYYY) |
| COD\_AUDITOR | varchar(10) |
| RESULTADO | Number(5,4) |
| TITULO\_CUESTIONARIO | varchar(50) |

*Tabla 2. Detalle del tipo de campo*

* 1. Cardinalidad

**Cardinalidad**: 32797 registros.



1. Análisis funcional y diagrama de arquitectura de flujo de datos
   1. Debemos hacer un diagrama de cómo se van a ir los datos

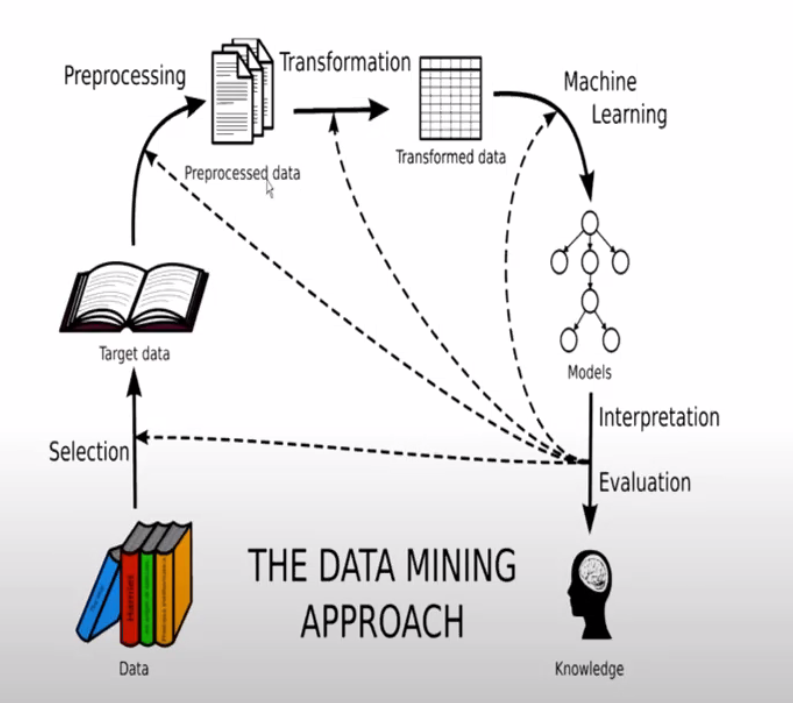
Extraer

Transformar

Cargar

**Diagrama de flujo de datos:**

**Ejemplo:**



1. **Generación de al menos 5 kpi´**

Tomando en consideración que el escenario plantea que el Departamento antifraude de la Compañía Mystey Shopping desea hacer un seguimiento y analizar la información relativa a las encuenstas que se realizan en los distintos centros de clientes, se describen los siguientes KPI´s, siendo la definición de estos como “*Indicadores Clave de Desempeño que se componen de variables, factores o unidades de medida que sirven para generar estrategias dentro de los departamentos de una empresa”*

a. Localidades en donde el resultado de encuesta/evaluación es mayor a 0.90 / tipo de cuestionario.

b. Tipo de cuestionario aplicado por localidad en el año 2014.

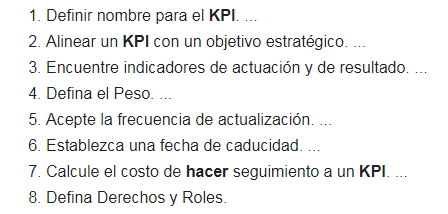
c. Población que se le aplicó a un determinado cuestionario en cada semestre del año 2013 y 2014.

d. Qué oficina aplicó cuestionarios y qué tipo en los años 2013 y 2014.

e. Qué auditores evaluaron encuestas por tipo cuestionario, localidad, provincia y población.

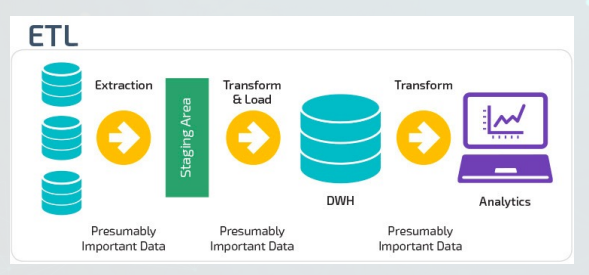
f. Proyecto evaluado por localidad / provincia en los años 2013 y 2014.

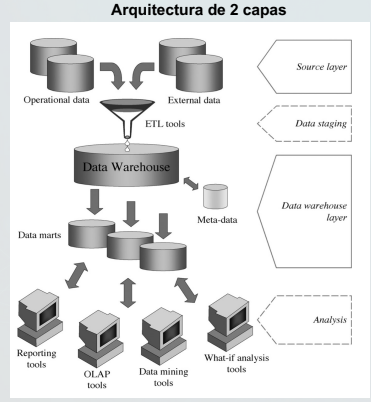
Nota: creo que lo que yo planteé como KPIs no está bien, nos tocaría ampliarle un poco más al KPI para que se pueda cumplir lo siguiente:



1. **Qué arquitectura de referencia se utilizaría?**

Arquitectura de referencia (Gustavo quedó en graficar la arquitectura en base a esta) habría que describir en palabras también.





1. Qué tecnología OLAP se utilizaría (debería ser ROLAP porque es pequeño, etc…)

Para este caso específico del diseño de la solución, describiremos a ambas tecnologías, OLAP utilizaría un servidor intermedio y en cambio ROLAP accede directamente a la base de datos relacional, por lo tanto, nuestra solución se basa en ROLAP que accederá a datos más directos, oportunos y actualizados que OLAP y también por tratarse de un escenario pequeño en dimensión.

La diferencia básica entre OLAP y ROLAP es su arquitectura, es decir, las diferentes formas en que OLAP y ROLAP obtienen y acumulan información en una base de datos relacional.

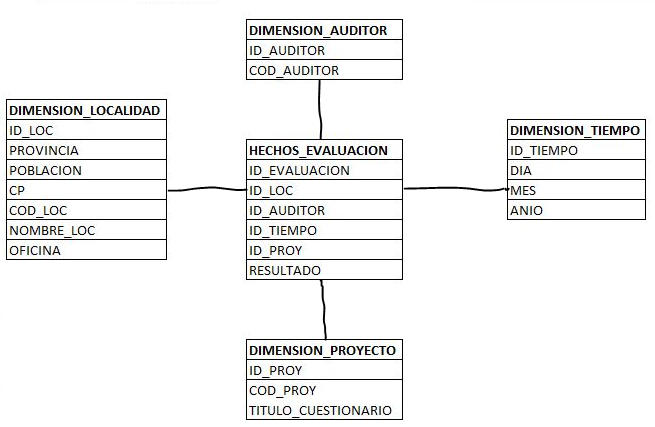
OLAP recoge los datos necesarios y los almacena en una memoria cache de datos especializada, esto permitiría un sólido rendimiento a un costo reducido. ROLAP ofrece una funcionalidad similar, pero realiza su análisis sobre la marcha, sin el paso intermedio de almacenar los datos en un servidor especial.

La ventaja radicaría en que permitirá a los usuarios realizar consultas rápidamente en bases de datos que son demasiado grandes para que un sistema OLAP pueda recorrerlas y analizarlas eficientemente.

1. Si se utiliza ROLAP ¿Cuál de estos dos modelos se ajustaría mejor, estrella o copo de nueve)

Para la solución se utilizará ROLAP (OLAP sobre relacional), ya que se va a usar una base de datos relacional para implementar la capa física y una capa de metadatos para implementar el modelo lógico multidimensional.

Se plantea un Modelo en **Estrella** para la solución, donde se tendrá una Tabla de Hechos denominada EVALUACION en el centro y sus puntas serán las dimensiones LOCALIDAD, AUDITOR, TIEMPO y PROYECTO.



Se plantea también una estrategia de diseño inicial de Kimball o bottom-up que permita comenzar a nivel local o departamental para escalar al resto de la organización:

1. Normalización.- nos va a explicar el sábado.
2. Si ya se tiene conectados las BD, y como están conectadas las tablas.
3. Si ya se tiene conectados las BD, y como están conectadas las tablas.
4. Si ya se tiene conectados las BD, y como están conectadas las tablas.

Hasta este punto 1 al 10 es conceptual, en papel

1. Se comenzaría a utiliza las herramientas. Aquí ya entra Pentaho
2. Se va usar la herramienta Power BI
3. Se va a ver mas adelante
4. Se va a realizar un dashboard,

Prueba edición ccabezas