

**UNIVERSIDAD DE EL SALVADOR**  
**FACULTAD DE INGENIERÍA Y ARQUITECTURA ESCUELA**  
**DE INGENIERÍA DE SISTEMAS INFORMÁTICOS**



**CURSO DE ESPECIALIZACIÓN DE INGENIERIA DE DATOS**

**ANALISIS DE INFORMACION SOBRE MANEJO DE COMPRAS DE LA  
EMPRESA CAPITAL GENERAL GESTIONADA A TRAVES DEL SISTEMA  
ERP NEXT**

PRESENTADO POR:

**CALZADILLA RAMOS, ANDREA MARIA**

**NAVAS HERNANDEZ, FRANCISCO JOSE**

**ROSALES MARTINEZ, MANUEL ANTONIO**

PARA OPTAR AL TÍTULO DE:

**INGENIERO DE SISTEMAS INFORMÁTICOS**

**Cuidad Universitaria, Diciembre de 2023**

**UNIVERSIDAD DE EL SALVADOR**

RECTOR:

**M. SC. JUAN ROSA QUINTANILLA**

SECRETARIO GENERAL:

**LIC. PEDRO ROSALIO ESCOBAR CASTANEDA**

**FACULTAD DE INGENIERÍA Y ARQUITECTURA**

DECANO:

**ING. LUIS SALVADOR BARRERA MANCIA**

SECRETARIO:

**ARQ. RAUL ALEXANDER FABIAN ORELLANA**

**ESCUELA DE INGENIERÍA DE SISTEMAS INFORMÁTICOS**

DIRECTOR:

**ING. CÉSAR AUGUSTO GONZALEZ RODRIGUEZ**

**UNIVERSIDAD DE EL SALVADOR.  
FACULTAD DE INGENIERÍA Y ARQUITECTURA.  
ESCUELA DE INGENIERÍA DE SISTEMAS INFORMÁTICOS.**

Curso de Especialización previo a la opción al Grado de:  
**INGENIERO DE SISTEMAS INFORMÁTICOS**

Título:

**ANALISIS DE INFORMACION SOBRE MANEJO DE COMPRAS DE LA  
EMPRESA CAPITAL GENERAL GESTIONADA A TRAVES DEL SISTEMA  
ERP NEXT**

Presentado por:

**CALZADILLA RAMOS, ANDREA MARIA  
NAVAS HERNANDEZ, FRANCISCO JOSE  
ROSALES MARTINEZ MANUEL ANTONIO**

Curso de Especialización Aprobado por:

Docente Asesor:

**ING. RENE FABRICIO QUINTANILLA GÓMEZ  
SAN SALVADOR, DICIEMBRE 2023**

Curso de Especialización Aprobado por:

Docente Asesor:

**Ing. Rene Fabricio Quintanilla Gómez**

# Índice

Introducción.....	7
Capítulo I: Especificación Del Proyecto.....	8
a) Situación Actual .....	8
I. Antecedentes .....	8
II. Descripción del problema .....	9
III. Planteamiento del problema.....	10
b) Objetivos.....	11
Objetivo General .....	11
Objetivos Específicos .....	11
c) Alcances .....	12
d) Justificación.....	13
e) Cronograma de Actividades.....	14
f) Presupuesto.....	15
Capítulo II: Análisis y Diseño de la propuesta de solución.....	17
a) Metodología de trabajo .....	17
Arquitectura de Kimball: Data Warehouse y Business Intelligence (DW/BI) .....	17
Sistemas Origen o Sistemas Fuente (Source Transactions .....	17
Sistemas ETL (Back Room) .....	17
Área de presentación (Front Room) .....	18
Aplicaciones de Inteligencia de Negocio (Business Intelligence) .....	18
b) Descripción de la propuesta de solución.....	18
Descripción del dataset y diccionario de datos.....	18
Resultado de la data profiling .....	26
Especificación de las necesidades analíticas .....	31
Modelo dimensional propuesto .....	31
Diagrama de estrella del modelo dimensional .....	32
Mapping de modelo dimensional y tablas del data set .....	33
c) Descripción de la tecnología a utilizar .....	40
d) Diagrama Arquitectónico de la solución .....	43
e) Descripción de cada componente de la solución.....	43
Capítulo III: Estrategia de Implementación de Propuesta de Solución .....	46
a) Estrategia de Implementación.....	46
Base de datos Mysql .....	46

Bucket S3.....	46
Redshift.....	47
Proceso ETL en Talend.....	47
Power BI.....	59
Plan de migración.....	61
Frecuencia.....	61
b) Presupuesto de Implementación (Distribuido en rubros) .....	61
c) Análisis de Resultados .....	63
Reporte de Monto de Compras.....	63
Reporte Cantidad de Productos por Proveedor.....	64
Reporte Cantidad de Productos Comprados .....	65
Reporte Monto por Producto y Proveedor.....	66
Reporte Cantidad de Compras.....	67
Conclusiones .....	68
Recomendaciones .....	69
Bibliografía .....	70
Anexos.....	71
Anexo 1: Diagrama de proceso de Orden de Compras.....	71
Anexo 2: Diagrama del Modelo Dimensional Data Warehouse.....	72
Anexo 3: Diagrama Gantt De Actividades .....	73

## Introducción

Actualmente el volumen de datos proveniente de distintas fuentes ha ido incrementando, por lo que al ser el principal recurso de la empresa es necesario la gestión correcta de los mismos. Es esencial la transformación adecuada de sus datos históricos, para tener la facilidad del análisis de una forma más intuitiva conllevando a toma de decisiones estratégicas para la mejora continua de los procesos operativos y empresariales.

Por tanto, se vuelve una necesidad para la empresa Capital General de mejorar el proceso de compras que se tiene dentro de la compañía, y para ello se plantea la creación de un DataWarehouse que ayudara al procesamiento de los datos y el análisis de los mismos. Se utiliza ERPNEXT para registrar las ordenes de compras, registrando los datos transaccionales e históricos relacionados al proceso de compras. Pero estos datos no tienen una transformación y un proceso que ayude a un análisis adecuado de este proceso dificultando la creación de informes que representen los datos de forma amigable para el área de compras.

Por lo que considerando todos estos factores se propone implementar el modelo dimensional haciendo uso de herramientas tales como ERPNext, Talend Open Studio, Amazon Web Services tales como: Amazon Redshift, Amazon S3 Y Amazon IAM. Estas tecnologías facilitaran la extracción, transformación y la carga de datos a nuestro sistema dimensional. Esto conllevara a la creación de áreas de almacenamientos en S3 para los datos que no han sido procesados (RAW), los datos que ya se encontraran procesados y que están listos para la carga (STAGE), y los están preparados para la información visual que en este caso serán los que estarán en la zona de PRESENTACION. Los datos requerirán tareas ETL para estar completamente completos y listo para cada una de estas zonas. Además, se implementará el uso de la base de datos en la nube en este caso Amazon Redshift ya que nos facilita su uso, gestión y almacenamiento. Brindando facilidad para la conexión con nuestra herramienta visual que será Power BI.

Es fundamental garantizar la seguridad en el desarrollo de la solución por lo cual se debe establecer el control de acceso a los datos en la nube por medio de Amazon IAM, estableciendo credenciales, permisos de accesos y políticas.

En términos generales, un Data Warehouse no es simplemente una infraestructura de almacenamiento de datos, sino una herramienta estratégica esencial para potenciar a las organizaciones en la era de la información, permitiéndoles convertir datos en decisiones y oportunidades en sus procesos empresariales.

# Capítulo I: Especificación Del Proyecto

## a) Situación Actual

### I. Antecedentes

La empresa Capital General es una compañía que ofrece soluciones tecnológicas y comercia principalmente productos de computadoras tales como CPU, tarjetas de videos, dispositivos de almacenamiento y mother board.

Por lo que los distintos proveedores con los cuales ellos realizan los procesos de compras para adquirir estos artículos se clasifican en estos 4 tipos.

Actualmente la empresa Capital General utiliza ERP NEXT para el registro y gestión de las compras, proveedores y los productos que deben comprar para el funcionamiento de la misma. Se pueden mencionar que algunos de los beneficios mas notables de el uso de este software es que ofrece una gestión sencilla, con una interfaz amigable para el uso diario de la empresa, facilitando la identificación de los productos y las ordenes de compras que se realizan.

Por lo que se puede describir la lógica de negocio implementada de la siguiente manera: Cuando ellos inician el proceso de una compra, se empieza por seleccionar el proveedor al cual solicitara los productos de la orden de compra y especifica la fecha requerida de la entrega.

Luego eligen los artículos que necesita pedir al proveedor, es buscado por el código del producto, establece la cantidad, los impuestos ya se encuentran predeterminados o si han establecido nuevos por ejemplo por costos adicionales como la distancia de donde se despacha hacia donde se suministra. Cabe destacar que el sistema brinda la facilidad que actualiza automáticamente el precio que tendrán de acuerdo a la cantidad y precios unitarios. Además, pueden existir ciertas condiciones que ellos establecen cuando el proveedor entregara el producto. El usuario guarda y envía la orden de compra.

Después de enviar la orden de compras, el departamento de compra tiene las opciones:

- Crear un recibo de compra: el cual se indica que se recibió los artículos y con ella se actualiza la información existente de la cantidad de ellos en el almacén, sin este recibo el sistema no permitirá actualizar las cantidades existentes.
- Crear la factura de compra: el cual es una opción que facilita la información que por ejemplo departamento de contabilidad necesita para llevar al día la contabilidad.
- Pago: El cual detalla el pago realizado en la orden de compra realizada, esta opción es generada a partir de la factura emitida

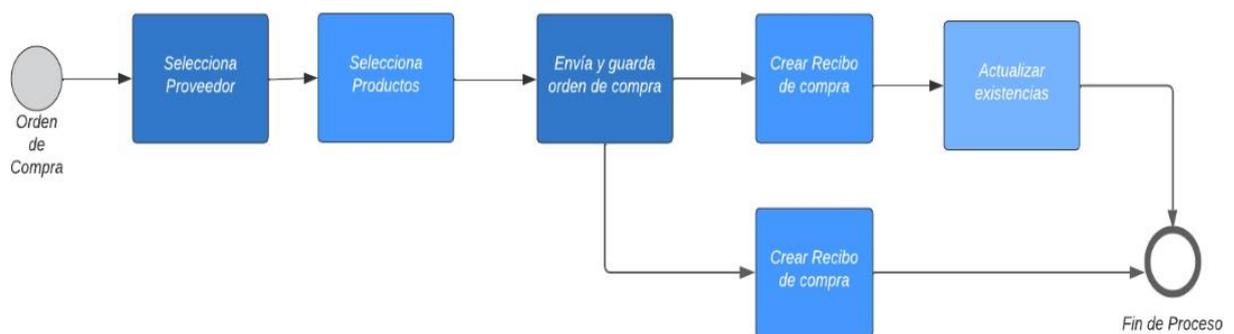


Imagen 1: Diagrama de Proceso de Orden de Compra

## **II. Descripción del problema**

Capital General compra diferentes artículos electrónicos los cuales los adquiere de diferentes proveedores, estos se solicitan por medio de las órdenes de compra. Sin embargo, al final de este proceso la empresa no tiene la información sintetizada para poder realizar el respectivo análisis de este departamento y como se lleva la gestión de compras.

Además, la velocidad a la que se generan y actualizan los datos, junto con la diversidad de formatos, agrega complejidad al proceso de toma de decisiones. Esta problemática no solo afecta la eficiencia operativa, sino que también obstaculiza la capacidad de la empresa para identificar oportunidades y mitigar riesgos de manera oportuna.

Presentando un problema a la hora de tomas de decisiones gerenciales específicamente en el área de compras por lo que es necesario conocer valores, este caso estos datos los hemos tomado como métricas para poder presentarlos en los reportes que apoyarán a este departamento, los cuales serán:

- Cantidad de compras por periodo de tiempo.
- Total, de productos por proveedor y en un periodo de tiempo
- Monto total de los productos
- Que producto fue comprado más y cual fue menos.

La implementación de un data warehouse emerge como una solución estratégica para superar estos desafíos. La creación de un repositorio centralizado permitirá la extracción, transformación y carga (ETL) de datos desde diversas fuentes hacia un entorno unificado. Esto facilitará análisis más rápidos y precisos, así como la generación de informes coherentes que respalden la toma de decisiones informada en todos los niveles de la organización.

### III. Planteamiento del problema

Con el objetivo de plantear el problema se describirá en función a partir del entradas, procesos y salidas que nos llevará a la implementación del datawarehouse.

Por lo que cada una de ellas se especifica a continuación:

- Entradas: Son las que se obtienen a partir del sistema ERP NEXT de las compras realizadas por la empresa Capital General.
- Procesos: Los distintos recursos que se utilizaron para la extracción, transformación y carga de datos tales como Amazon S3, OpenTalent, Amazon Red shift asimismo como los que servirán para la presentación, visualización de datos y creación de informes con estos valores en este caso Power BI.
- Salidas: El dashboard con los distintos reportes que se obtendrán de la transformación de datos tales como la cantidad de productos comprados en una fecha, el monto de compras, productos más comprados y cuáles no.



Imagen 2: Digrama de caja negra que representa la situación actual de la empresa (Estado A) y la situación con la propuesta de solución (Estado B)

## **b) Objetivos**

### **Objetivo General**

Construir una solución de análisis empresarial a nivel de data warehouse para el área de compras de la empresa Capital General que facilite y apoye en la toma de decisiones a nivel gerencial.

### **Objetivos Específicos**

- Analizar la situación actual de la empresa Capital General para el desarrollo de la solución propuesta.
- Extraer la base de datos del módulo de compras del sistema ERP(ERPNext) utilizado por la empresa Capital General a través de un Sistema Gestor de Base de Datos para el análisis de la información.
- Diseñar procesos ETL, que permitan extraer, transformar y cargar datos que provienen del sistema fuente ERP.
- Implementar el servicio de almacenamiento de objetos de S3 de AWS para la disponibilidad, almacenamiento y transformación de los datos desde el sistema transaccional.
- Diseñar e Implementar la arquitectura de data warehouse mediante un modelo dimensional de estrella utilizando el servicio de Amazon Redshift de AWS que soporte la carga de datos proveniente de S3
- Diseñar Dashboard(Gráficos) que presenten información optimizada limpia y de calidad para la toma de decisiones gerenciales.

## **c) Alcances**

### **Extracción de Datos:**

- Identificar y extraer datos relevantes del ERP Next, centrándose en el módulo de órdenes de compra.
- Incluir información sobre proveedores, productos, órdenes de compra, recibos de compra.
- Transformación y Limpieza de Datos:
- Normalizar y limpiar los datos extraídos para garantizar consistencia.
- Convertir fechas, formatos y unidades según sea necesario.
- Manejar datos faltantes y valores atípicos.

### **Modelado de Datos:**

- Diseñar un modelo de datos adecuado para el data warehouse, teniendo en cuenta la relación entre las entidades (proveedores, productos, órdenes, recibos).
- Definir claves primarias y foráneas.
- Carga de Datos en AWS Redshift:
- Utilizar Talend para cargar los datos transformados en la base de datos Redshift.
- Implementar procesos ETL para mantener la consistencia y actualizar los datos.

### **Análisis de Datos:**

- Proporcionar capacidades analíticas para que los usuarios puedan realizar consultas y obtener información valiosa.
- Crear informes y paneles de control para facilitar la toma de decisiones.

#### **d) Justificación**

La obtención de datos desde un sistema transaccional de manera específica era una forma común para la creación a futuro de reportes de los cuales las empresas se les facilitaría el análisis para futuras decisiones o acciones de las cuales dependerán la mejora de las misma.

Sin embargo, cada vez el procesamiento de grandes cantidades de datos se vuelve complejo para sistemas transaccionales, aun métodos que ayudan a trabajar con datos no estructurados a cierto punto se vuelven pocos eficientes.

Una manera de optimizar y mejorar estas deficiencias es el uso de Datawarehouse, la cual nos permite gestionar diferentes orígenes de datos incluso si estos no son estructurados, facilitando el manejo de grandes volúmenes de datos conllevando a que nuestro usuario final tenga accesibilidad a ellos, favoreciendo las tareas analíticas lo cual es fundamental en cada negocio.

El proceso de compras en una empresa es una tarea fundamental del negocio por lo que facilitar el estudio de las variables que interesan a una compañía para futuras proyecciones o análisis de las conveniencias de ciertos productos al crecimiento de ellas es una de las principales facilidades que el datawarehouse ayudara. La presentación de estos datos de una forma que el usuario será capaz de extraer y transformar para analizarlos a través de reportes de herramientas que ayudan a presentar datos de una forma amigable y entendible para las distintas áreas de interés de la empresa.

## e) Cronograma de Actividades

Name	Begin date	End date
Introducción a la lógica del Negocio	16/6/23	17/6/23
Análisis Data Set y Base de Datos Transaccional	19/6/23	20/6/23
Data Profiling	20/6/23	21/6/23
Definición Métricas	21/6/23	21/6/23
Diseño Dimensional Data Warehouse	22/6/23	24/6/23
Entrega Primera Etapa	29/6/23	29/6/23
Instalación Talend Open Studio	7/7/23	9/7/23
Aprendizaje AWS	10/8/23	7/9/23
Creación Bucket en S3	23/8/23	25/8/23
Llenado de Datos	20/9/23	3/11/23
Configuración IAM	22/9/23	30/9/23
Creación Base en Amazon Redshift	5/10/23	13/10/23
Ejecución de y Configuración de Jobs en Open Talend	27/9/23	21/11/23
Aprendizaje Power BI	20/10/23	25/11/23
Creación Reporte Power BI	9/12/23	9/12/23
Entrega Segunda Etapa	9/12/23	9/12/23

Imagen 3: Listado de actividades que comprenden la Primera y Segunda etapa de entrega

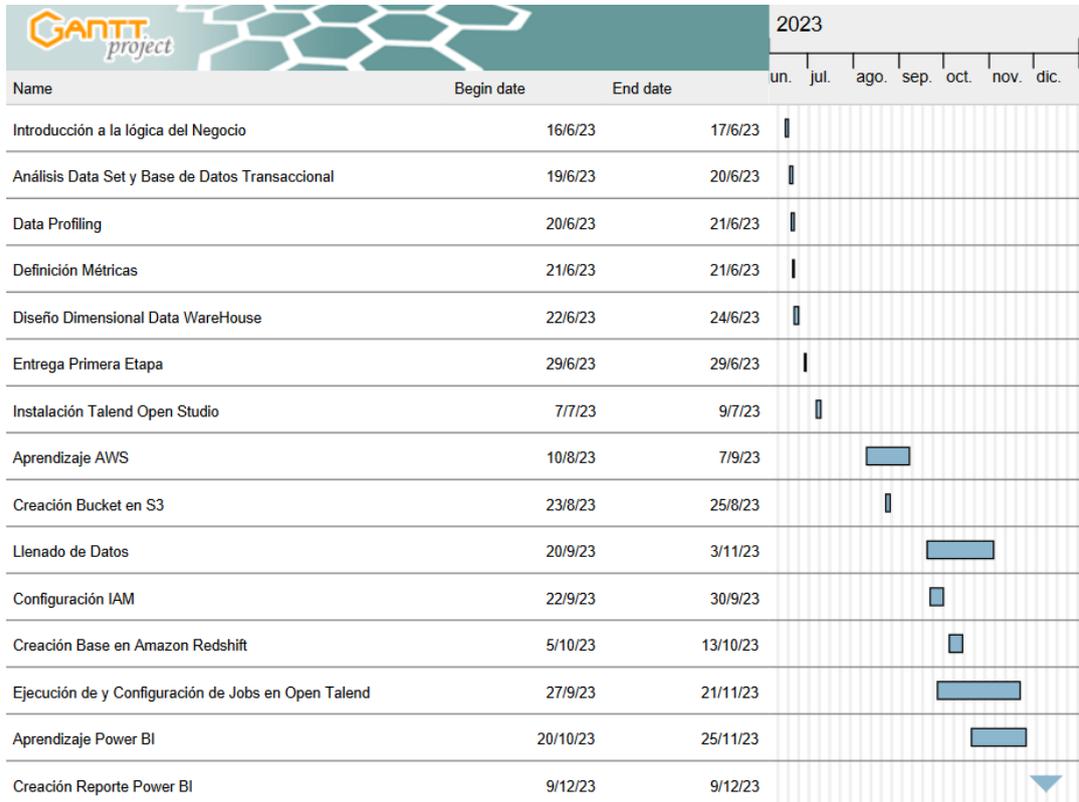


Imagen 4: Diagrama de Gantt-Cronograma de actividades

## f) Presupuesto

### Presupuesto de Desarrollo:

En esta sección se presentan los costos incurridos durante el desarrollo del proyecto, se consideran recursos humanos, tecnológicos y servicios básicos que fueron necesarios para el desarrollo de la solución.

### Recurso humano:

Para cada integrante del equipo de desarrollo se toma como base un salario mínimo de \$1200 mensuales a razón de sus servicios como ingeniero de datos:

Recurso Humano	Cantidad	Salario Mensual (\$)	Duración del Proyecto (Meses)	Salario Total (\$)
Ingeniero de datos	3	\$1200	6	\$21,600

Tabla 1: Presupuesto de desarrollo - Recurso Humano

### Servicios básicos:

Recurso	Tiempo Estimado (Meses)	Monto Mensual (\$)	Total (\$) x Integrante	Total (\$)
Energia Electrica	6	\$25	\$150	\$450
Internet	6	\$35	\$210	\$630
<b>Total Servicios básicos</b>				<b>\$1080</b>

Tabla 2: Presupuesto de desarrollo - Servicios básicos

### Costos AWS:

Se consideran los costos representativos por los servicios de aws usados en el desarrollo de la solución. Los precios estimados mensuales se obtienen mediante AWS Pricing Calculator para los servicios de S3 y Redshift.

### Amazon S3

Tiered price for: 10 GB  
10 GB x 0,023 USD = 0,23 USD  
Costo total de la capa = 0,23 USD (coste de almacenamiento en S3 Estándar)  
4000 Solicitudes PUT para almacenamiento de S3 Standard x 0,000005 USD por solicitud = 0,02 USD (coste de solicitudes PUT en S3 Estándar)  
4000 Solicitudes GET en un mes x 0,000004 USD por solicitud = 0,0016 USD (coste de solicitudes GET en S3 Estándar)  
2 GB x 0,0007 USD = 0,0014 USD (coste por las devoluciones de S3 Select)  
5 GB x 0,002 USD = 0,01 USD (coste de escaneados en S3 Select)  
0,23 USD + 0,0016 USD + 0,02 USD + 0,0014 USD + 0,01 USD = 0,26 USD (Total de almacenamiento de S3 Standard, solicitudes de datos, coste de S3 Select)  
**Coste de S3 Estándar (mensual): 0.26 USD**

Redondear al valor unitario superior (2.000000) = 2 partes  
**Coste de S3 Estándar (inicial): 0.00 USD**

Costo inicial total: 0.00 USD  
Costo total mensual: 0,26 USD

Mostrar detalles ▼ Cancelar

Imagen 5: Presupuesto de desarrollo - Amazon S3

El costo mensual para el servicio de S3 es de \$0.26.

## Redshift

Para el desarrollo del proyecto y el uso del servicio de redshift, se estima el siguiente costo considerando adquirir una instancia del tipo dc2.large, con 2 CPU virtuales, 1 nodo y una memoria de 15GiB.

<b>1 en ejecución x 0.25 USD por hora x 240 horas en un mes = 60.0000 USD</b>
<b>Costo de la instancia de Redshift (mensual): 60.00 USD</b>
<b>Costo de la instancia de Redshift (inicial): 0.00 USD</b>

*Imagen 6: Presupuesto de desarrollo - Amazon Redshift*

El costo mensual para el servicio de Amazon Redshift es de \$60.

Es importante mencionar que los servicios de Amazon S3 y Amazon Redshift contempla 3 meses, tiempo en el cual se hizo uso de los servicios en la nube para la etapa de desarrollo, y el monto total se ve reflejado en la tabla de presupuesto.

## Costo Total de desarrollo

La siguiente tabla presenta el monto del costo total de desarrollo del proyecto para un periodo de 6 meses de duración del proyecto, considerando los costos de los recursos previamente calculados.

<b>Recurso</b>	<b>Total (\$) Duración del proyecto</b>
<b>Humano</b>	21,600
<b>Servicios Básicos</b>	1,080
<b>Amazon S3</b>	0.78
<b>Amazon Redshift</b>	180
<b>Total</b>	<b>22,860.78</b>

*Tabla 3: Costo Total - Presupuesto de desarrollo*

## Capítulo II: Análisis y Diseño de la propuesta de solución

### a) Metodología de trabajo

Para el desarrollo de este proyecto y la construcción del Data Warehouse se implementó la metodología de Ralph Kimball, la cual es la metodología más utilizada para la construcción de soluciones con Data Warehouse.

A continuación, se describen los componentes que conforman la metodología de Ralph Kimball y que deben ser considerados en un entorno DW/BI.

### Arquitectura de Kimball: Data Warehouse y Business Intelligence (DW/BI)

Existen cuatro componentes que conforman esta arquitectura:

- Sistemas Origines
- Sistemas ETL
- Área de presentación de datos
- Aplicaciones BI

La siguiente imagen muestra gráficamente la arquitectura de Kimball junto a sus elementos centrales.

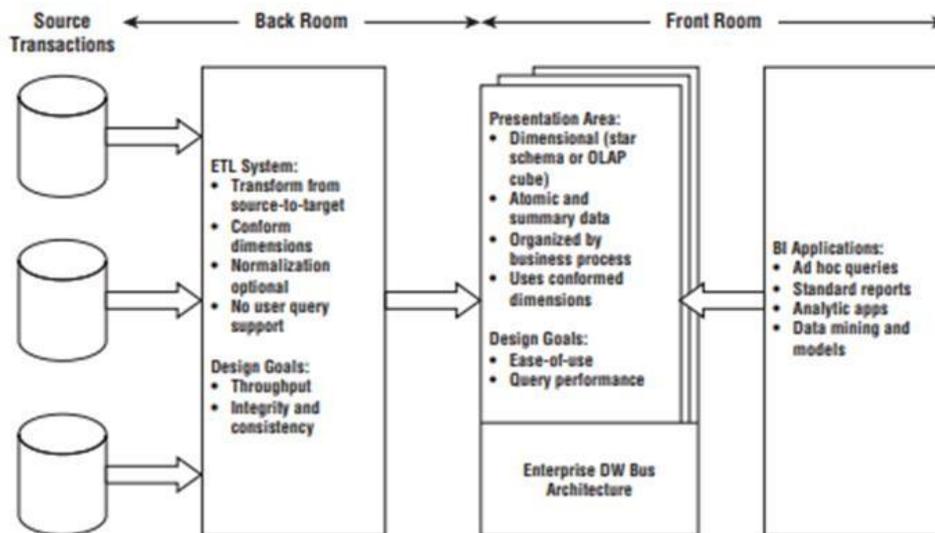


Imagen 7: Diagrama de arquitectura de Kimball

### Sistemas Origen o Sistemas Fuente (Source Transactions)

Son sistemas y/o aplicaciones que capturan las transacciones del negocio, es decir donde se almacenan y procesan los datos de origen operativo de la empresa, algunos ejemplos de estos sistemas pueden ser: Sistema de ventas, contabilidad, inventario, software ERP, entre otros.

Para el caso particular de este proyecto, la empresa Capital General cuenta con el software ERP Next donde se lleva el registro de su operatividad.

### Sistemas ETL (Back Room)

Es la capa de procesamiento y almacenamiento de datos del entorno de DW/BI, y cumplen la función de extracción, transformación y carga (ETL, por sus siglas en inglés) de los datos.



## Diccionarios de Datos:

**tabpurchase taxes and charges:** Tasas e impuestos de compra

Columna	Tipo de dato	Descripción	Key	NULL	Restricciones	Notas
<b>name</b>	Varchar (140)	Nombre del impuesto.	Primary Key	No		Llave primaria
<b>creation</b>	datetime	Fecha de creación		Si		
<b>modified</b>	datetime	Fecha de modificación		Si		
<b>Modified_by</b>	Varchar (140)	Quien modifíco		Si		
<b>owner</b>	Varchar (140)	dueño		Si		
<b>docstatus</b>	Int (1)	Estado del registro		no		
<b>idx</b>	Int (8)	Correlativo generado		No		
<b>category</b>	Varchar (140)	Categoría del impuesto		Si		
<b>add_deduct_tax</b>	Varchar (140)			Si		
<b>charge_type</b>	Varchar (140)			Si		
<b>included_in_print_rate</b>	Int (1)	Incluir en la impresión		No	Default = 0	
<b>account_head</b>	Varchar (140)	Encabezado de cuenta		si		
<b>description</b>	Varchar (140)	Descripción del elemento		si		
<b>rate</b>	Decimal (21,9)	Calificación		no	Default = 0	
<b>cost_center</b>	Varchar (140)	Centro de costo		Si		
<b>account_currency</b>	Varchar(140)			Si		
<b>tax_amount</b>	decimal (21,9)			No		
<b>tax_amount_after_discount_amount</b>	decimal (21,9)			No		
<b>total</b>	decimal (21,9)			No		
<b>base_tax_amount</b>	decimal (21,9)			No		
<b>parent</b>	varchar (140)	Si tiene padre		Si		
<b>parentfield</b>	varchar (140)	Padre ID		Si	index	
<b>parenttype</b>	varchar(140)	Tipo de padre		Si		

Tabla 4: Campos de tabla Impuestos de compras

## Indices

Nombre de la llave	Tipo	NUL L	Columna	Comentario
<b>PRIMARY</b>	PRIMARY	Si	name	Llave primaria
<b>parent</b>	Index	No	parent	Index
<b>modified</b>	Index	no	modified	index

Tabla 5: Indices de la tabla Impuestos de compras

**Cotejamiento:** utf8mb4\_unicode\_ci

**tabpurchase order:** Orden de compra

Columna	Tipo de dato	Descripción	Key	NULL	Restricciones	Notas
<b>name</b>	Varchar (140)	Nombre del impuesto.	Primary Key	No		Llave primaria
<b>creation</b>	datetime	Fecha de creación		Si		
<b>modified</b>	datetime	Fecha de modificación		Si		
<b>owner</b>	Varchar (140)	dueño		Si		
<b>docstatus</b>	Int (1)	Estado del registro		no		
<b>idx</b>	Int (8)	Correlativo generado		No		
<b>Title</b>	Varchar(140)	Titulo		Si	Default: '{supplier_name}'	
<b>naming_series</b>	Varchar(140)			Si		
<b>supplier</b>	Varchar(140)	Id del proveedor	index	si		
<b>supplier_name</b>	Varchar(140)	Nombre del proveedor		Si		
<b>apply_tds</b>	Int(1)			No		
<b>tax_withholding_category</b>	Varchar(140)			si		
<b>is_subcontracted</b>	Int(1)	Subcontrato		no		
<b>supplier_warehouse</b>	Varchar(140)	Bodega del proveedor		Si		
<b>company</b>	Varchar(140)	Empresa		Si		
<b>transaction_date</b>	Date	Fecha de la transacción		Si		
<b>schedule_date</b>	Date			Si		
<b>order_confirmation_no</b>	Varchar(140)	Numero de orden de confirmación		Si		
<b>order_confirmation_date</b>	Date	Fecha de orden de confirmación		Si		

<b>cost_center</b>	Varchar(140)	Centro de costo		Si		
<b>project</b>	Varchar(140)	Proyecto		Si		
<b>Customer</b>	Varchar(140)	Cliente		Si		
<b>customer_name</b>	Varchar(140)	Nombre del cliente		Si		
<b>customer_contact_person</b>	Varchar(140)	Contacto del cliente		Si		
<b>customer_contact_display</b>	Text			Si		
<b>customer_contact_mobile</b>	Text	Celular del cliente		Si		
<b>customer_contact_email</b>	Longtext	Correo del cliente		Si		
<b>supplier_address</b>	Varchar(140)	Dirección del proveedor		Si		
<b>address_display</b>	Varchar(140)	Dirección		SI		
<b>contact_person</b>	Varchar(140)	Contacto de la persona		Si		
<b>contact_display</b>	Varchar(140)			Si		
<b>shipping_address</b>	Varchar(140)	Dirección de envío		Si		
<b>shipping_address_display</b>	Text			Si		
<b>billing_address</b>	Varchar(140)	Dirección de pago		Si		
<b>Currency</b>	Varchar(140)	Divisa		Si		
<b>total_qty</b>	Decimal(21,9)	Total de cantidad		No		
<b>Base_total</b>	Decimal(21,9)			No		
<b>Total</b>	Decimal(21,9)	Total		No		
<b>set_reserve_warehouse</b>	Varchar(140)	Asignar bodega resevada		Si		
<b>tax_category</b>	Varchar(140)	Categoría del impuesto		Si		
<b>shipping_rule</b>	Varchar(140)	Reglas de envío		No		
<b>taxes_and_charges</b>	Decimal(21,9)	Impuestos y cargos		No	Default=0	
<b>other_charges_calculation</b>	Decimal(21,9)	Otros cargos calculados		No	Default=0	
<b>taxes_and_charges_added</b>	Decimal(21,9)	Impuestos y cargos agregados		No	Default=0	
<b>taxes_and_charges_deducted</b>	Decimal(21,9)	Impuestos y cargos deducidos		No	Default=0	

<b>total_taxes_and_charges</b>	Decimal(21,9)	Total, de impuestos y cargos		No	Default=0	
<b>additional_discount_percentage</b>	Decimal(21,9)	Adicional porcentaje de descuento		no	Default=0	
<b>apply_discount_on</b>	varchar(140)			No	Default='Grand Total'	
<b>'Grand Total'</b>	Decimal(21,9)	Total redondeado		no	Default=0	
<b>disable_rounded_total</b>	Int(1)	Deshabilitar redondeado		No	Default=0	
<b>advance_paid</b>	Decimal(21,9)	Pagos avanzados		No	Default=0	
<b>payment_terms_template</b>	Varchar(140)	Términos de pago plantilla		Si		
<b>Status</b>	Varchar(140)	Estado		Si	Default='Draft'	
<b>per_billed</b>	Decimal(21,9)	Por facturado		No	Default=0	
<b>Per_received</b>	Decimal(21,9)	Por recibir		No	Default=0	
<b>Terms</b>	Longtext	Términos		Si		
<b>Language</b>	Varchar(140)	Lenguaje		Si		
<b>group_same_items</b>	Int(1)			No		
<b>from_date</b>	Date			Si		
<b>To_date</b>	Date			Si		
<b>represents_company</b>	Varchar(140)	Representante de la compañía		Si		
<b>inter_company_order_reference</b>	Varchar(140)	Referencia de la orden de la empresa		Si		
<b>_user_tags</b>	Text	Etiquetas de usuario		si		
<b>_comments</b>	Text	Comentario		Si		

Tabla 6: Campos de la tabla Orden de compra

## Indices

Nombre de la llave	Tipo	NUL	Columna	Comentario
<b>PRIMARY</b>	PRIMARY	Si	name	Llave primaria, Unico
<b>supplier</b>	Index	No	supplier	Index
<b>transaction_date</b>	Index	No	transaction_date	Index
<b>status</b>	Index	No	status	Index
<b>modified</b>	Index	no	modified	Index

Tabla 7: Índices de la tabla Orden de compra

Cotejamiento: utf8mb4\_unicode\_ci

**tabpurchase order item:** artículos de orden de compra

Columna	Tipo de dato	Descripción	Key	NULL	Restricciones	Notas
<b>name</b>	Varchar (140)	Nombre del impuesto.	Primary Key	No		Llave primaria
<b>creation</b>	datetime	Fecha de creación		Si		
<b>modified</b>	datetime	Fecha de modificación		Si		
<b>owner</b>	Varchar (140)	dueño		Si		
<b>docstatus</b>	Int (1)	Estado del registro		no		
<b>idx</b>	Int (8)	Correlativo generado		No		
<b>Item_code</b>	Varchar(140)	Código del item		Si		
<b>Brand</b>	Varchar(140)			Si		
<b>product_bundle</b>	Varchar(140)	Paquete del producto		Si		
<b>fg_item</b>	Varchar(140)			Si		
<b>fg_item_qty</b>	Decimal(21,9)			No	Default=1	
<b>schedule_date</b>	Date	Fecha de programación		Si		
<b>expected_delivery_date</b>	Date	Fecha de entrega estimada		Si		
<b>item_group</b>	Varchar(140)	Item de grupo		Si		
<b>Description</b>	Text			Si		
<b>Image</b>	Text	Imagen		Si		
<b>qty</b>	Decimal(21,9)	Cantidad		No	Default=0	
<b>Stock_uom</b>	Varchar(140)			Si		
<b>conversion_factor</b>	Varchar(140)	Factor de conversión		No	Default = 0	
<b>stock_qty</b>	Decimal(21,9)	Cantidad en stock		Si	Default=0	
<b>price_list_rate</b>	Decimal(21,9)	Precio de lista		Si	Default=0	
<b>margin_type</b>	Varchar(140)	Tipo de margen		No		
<b>rate_with_margin</b>	Decimal(21,9)	Tarifa de margen		No		
<b>discount_percentage</b>	Decimal(21,9)	Descuento en porcentaje)		Si	Default=0	
<b>discount_amount</b>	Decimal(21,9)	Importe de descuento		Si	Default=0	
<b>base_rate_with_margin</b>	Decimal(21,9)	Tarifa base con margen		Si	Default=0	
<b>Rate</b>	Decimal(21,9)	Tasa		Si	Default=0	
<b>Amount</b>	Decimal(21,9)			Si	Default=0	
<b>item_tax_template</b>	Varchar(140)			No		

<b>base_rate</b>	Decimal(21,9)	Tasa base		Si	Default=0	
<b>base_amount</b>	Decimal(21,9)			Si	Default=0	
<b>pricing_rules</b>	Text	Precio de reglas		No		
<b>is_free_item</b>	Int(1)	Es gratis el item		Si	Default=false	
<b>Net_rate</b>	Decimal(21,9)	Tasa neta		No	Default=0	
<b>net_amount</b>	Decimal(21,9)	Importe neto			Default=0	
<b>base_net_rate</b>	Decimal(21,9)	Tasa neta base		No	Default = 0	
<b>base_net_amount</b>	Decimal(21,9)	Importe neto base		No	Default=0	
<b>from_warehouse</b>	Varchar(140)	De que bodega		Si		
<b>Warehouse</b>	Varchar(140)	Bodega		Si		
<b>actual_qty</b>	Decimal(21,9)	Cantidad actual		No	Default= 0	
<b>company_total_stock</b>	Decimal(21,9)			No	Default = 0	
<b>material_request</b>	Varchar(140)			Si		
<b>material_request_item</b>	Varchar(140)			Si		
<b>sales_order</b>	Varchar(140)	Orden de venta		Si		
<b>sales_order_item</b>	Varchar(140)	Orden de venta item		Si		
<b>sales_order_packed_item</b>	Varchar(140)	Articulo de pedido de venta		Si		
<b>supplier_quotation</b>	Varchar(140)	Oferta del proveedor		Si		
<b>supplier_quotation_item</b>	Varchar(140)	Presupuesto del articulo del proveedor		Si		
<b>delivered_by_supplier</b>	Varchar(140)	Entregado por proveedor		Si		
<b>received_qty</b>	Decimal(21,9)	Cantidad recibida		No	Default=0	
<b>returned_qty</b>	Decimal(21,9)	Cantidad retornada		No	Default=0	
<b>expense_account</b>	Varchar(140)	Cuenta de gastos		Si		
<b>Manufacturer</b>	Varchar(140)	Fabricante		Si		
<b>manufacturer_part_no</b>	Varchar(140)	Numero del fabricante		Si		
<b>include_explored_items</b>	Int(1)	Elementos explotados		Si		
<b>weight_per_unit</b>	Decimal(21,9)	Peso por unidad		No	Default=0	
<b>total_weight</b>	Decimal(21,9)	Total de peso		No	Default=0	
<b>Project</b>	Varchar(140)	Proyecto		Si		
<b>cost_center</b>	Varchar(140)					
<b>is_fixed_asset</b>	Int(1)	Es activo fijo		Si		

<b>item_tax_rate</b>	Decimal(21,9)	Tasa de impuesto del articulo		Si		
<b>production_plan</b>	Varchar(140)	Plan de producción		Si		
<b>Parent</b>	Varchar(140)	Padre		Si		

Tabla 8: Campos de la tabla Articulos de orden de compra

## Indices

Nombre de la llave	Tipo	NUL L	Columna	Comentario
<b>PRIMARY</b>	PRIMARY	No	name	Llave primaria
<b>modified</b>	Index	Si	modified	Index
<b>item_code</b>	Index	Si	item_code	
<b>expected_delivery_date</b>	Index	Si	expected_delivery_date	
<b>material_request</b>	Index	Si	material_request	
<b>material_request_item</b>	Index	Si	material_request_item	
<b>sales_order</b>	Index	Si	sales_order	
<b>sales_order_item</b>	Index	Si	sales_order_item	
<b>Parent</b>	Index	Si	Parent	
<b>Modified</b>	Index	Si	Modified	
<b>item_code_warehouse_index</b>	Index	si	warehouse	

Tabla 9: Indices de la tabla Artículo de orden de compra

**Cotejamiento:** utf8mb4\_unicode\_ci

**tabsupplier:** proveedor

Columna	Tipo de dato	Descripción	Key	NULL	Restricciones	Notas
<b>name</b>	Varchar (140)	Nombre del impuesto.	Primary Key	No		Llave primaria
<b>creation</b>	datetime	Fecha de creación		Si		
<b>modified</b>	datetime	Fecha de modificación		Si		
<b>owner</b>	Varchar (140)	dueño		Si		
<b>docstatus</b>	Int (1)	Estado del registro		no		
<b>idx</b>	Int (8)	Correlativo generado		No		
<b>naming_series</b>	Varchar (140)	Nombre de serie		Si		
<b>supplier_name</b>	Varchar (140)	Nombre del proveedor		Si		
<b>Country</b>	Varchar (140)	País		Si		

<b>default_bank_account</b>	Varchar (140)	Cuenta de banco		Si		
<b>tax_id</b>	Varchar (140)	Impuesto		Si		
<b>tax_category</b>	Varchar (140)	Categoría		Si		
<b>Image</b>	Text	Imagen		Si		
<b>supplier_group</b>	Varchar (140)	Grupo de proveedor		Si		
<b>supplier_type</b>	Varchar (140)	Tipo de proveedor		Si		
<b>is_internal_supplier</b>	Int (1)	Es proveedor interno		no	Default=0	
<b>disabled</b>	Int (1)	Deshabilitado		Si		
<b>is_transporter</b>	Int (1)	Es transportador		Si		
<b>represents_company</b>	Varchar (140)	Representante de la empresa		Si		
<b>default_currency</b>	Varchar (140)	Moneda predeterminada		Si		
<b>default_price_list</b>	Varchar (140)	Lista de precios		Si		
<b>supplier_primary_contact</b>	Varchar (140)	Contacto del proveedor		Si		
<b>mobile_no</b>	Varchar (140)	Numero de teléfono		Si		
<b>email_id</b>	Varchar (140)	Correo electrónico		Si		
<b>supplier_primary_address</b>	Varchar (140)	Dirección primaria del proveedor		Si		
<b>payment_terms</b>	Varchar (140)	Términos de pago		si		
<b>Website</b>	Varchar (140)	Sitio web				

Tabla 10: Campos de la tabla Proveedor

## Indices

Nombre de la llave	Tipo	NUL	Columna	Comentario
<b>PRIMARY</b>	PRIMARY	no	name	Llave primaria
<b>modified</b>	Index	Si	modified	Index

Tabla 11: Indices de la tabla de Proveedor

## Resultado de la data profiling

El data profiling o perfilado de datos permite descubrir, comprender y organizar datos mediante la identificación de sus características y la evaluación de su calidad. Este proceso puede revelar si los datos están completos o si son únicos, si se detectan errores y patrones inusuales y si se determina la facilidad de uso.

Herramientas utilizadas para realizar el Data Profiling

## DataCleanner

Para realizar el perfilado de datos se utilizó DataCleaner la cual es una herramienta que sirve para analizar la calidad de los datos obtenidos, con capacidad para encontrar patrones y supervisar los valores de los datos.

## Resultado del Data Profiling

Utilizando la herramienta antes mencionada y filtrando solo aquellas columnas de las tablas del módulo de compras establecido anteriormente, se obtuvieron los siguientes resultados: **(tener en cuenta que se ha mapeado solo las que se utilizaran para desarrollar el datawarehouse)**

Supplier	
Los campos de interés de esta tabla para el análisis correspondiente de este proyecto son: <ul style="list-style-type: none"><li>• name (PK)</li><li>• supplier_name</li><li>• supplier_group</li></ul>	
Campo	Observación
name(PK)	Cuando se guarda un proveedor el sistema genera una identidad o nombre único. <ul style="list-style-type: none"><li>• No existen valores nulos o en blanco.</li></ul>
supplier_name	Nombre con el cual se guarda el proveedor y se puede usar para referir al proveedor en varias transacciones de compra <ul style="list-style-type: none"><li>• No existen valores nulos o en blanco.</li><li>• Los nombres deben convertirse a mayúsculas mediante un proceso ETL.</li></ul>
supplier_group	Categoría a la que pertenece un proveedor, se configura en el sistema al crearse un nuevo registro de proveedor. <ul style="list-style-type: none"><li>• Existen valores nulos o en blanco.</li></ul>

Tabla 12: Data Profiling de la tabla Proveedor

Item
Los campos de interes de esta tabla para el análisis correspondiente de este proyecto son: <ul style="list-style-type: none"><li>• name(PK)</li><li>• item_code(UK)</li><li>• item_name</li><li>• item_group</li></ul>

<b>Campo</b>	<b>Observación</b>
<b>name(PK)</b>	Identificador único, se genera al registrar un producto. <ul style="list-style-type: none"> <li>No existen valores nulos o en blanco.</li> </ul>
<b>item_code</b>	Código del producto o Item. <ul style="list-style-type: none"> <li>No existen valores nulos o en blanco.</li> </ul>
<b>item_name</b>	Nombre con el cual se registra el producto. <ul style="list-style-type: none"> <li>No existen valores nulos o en blanco.</li> <li>Los nombres deben convertirse a mayúsculas mediante un proceso ETL.</li> </ul>
<b>item_group</b>	Categoría o grupo a la que pertenece el producto, entre los valores están: <ul style="list-style-type: none"> <li>All Item Groups</li> <li>Consumable</li> <li>Products</li> <li>Raw Material</li> <li>Services</li> <li>Sub Assemblies</li> </ul>

Tabla 13: Data profiling de la tabla Producto

<b>Item Price</b>	
<b>Los campos de interés de esta tabla para el análisis correspondiente de este proyecto son:</b> <ul style="list-style-type: none"> <li>name(PK)</li> <li>item_code(FK)</li> <li>price_list_rate</li> </ul>	
<b>Campo</b>	<b>Observación</b>
<b>name(PK)</b>	Identificador único, al generarse un nuevo registro. <ul style="list-style-type: none"> <li>No existen valores nulos o en blanco.</li> </ul>
<b>item_code(FK)</b>	Hace referencia al código del producto o Item. <ul style="list-style-type: none"> <li>No existen valores nulos o en blanco.</li> </ul>
<b>price_list_rate</b>	Lista de precios por producto. <ul style="list-style-type: none"> <li>No existen valores nulos o en blanco.</li> </ul>

Tabla 14: Data profiling de la tabla Precio producto

<b>Purchase Order</b>
<b>Los campos de interés de esta tabla para el análisis correspondiente de este proyecto son:</b> <ul style="list-style-type: none"> <li>name(PK)</li> <li>creation</li> </ul>

<ul style="list-style-type: none"> <li>• <b>modified</b></li> <li>• <b>supplier</b></li> <li>• <b>total_qty</b></li> <li>• <b>base_total</b></li> <li>• <b>total</b></li> <li>• <b>total_taxes_and_charges</b></li> <li>• <b>grand_total</b></li> <li>• <b>discount_amount</b></li> <li>• <b>schedule_date</b></li> <li>• <b>order_confirmation_date</b></li> </ul>	
<b>Campo</b>	<b>Observación</b>
<b>name(PK)</b>	Identificador único, al generarse una nueva orden de compra. <ul style="list-style-type: none"> <li>• No existen valores nulos o en blanco.</li> </ul>
<b>creation</b>	Fecha de creación de la orden de compra.
<b>modified</b>	Fecha de modificación de la orden de compra.
<b>supplier</b>	Identificador del proveedor. <ul style="list-style-type: none"> <li>• No existen valores nulos o en blanco.</li> </ul>
<b>total_qty</b>	Identifica la cantidad total de Items.
<b>base_total</b>	Monto total base <ul style="list-style-type: none"> <li>• No existen valores nulos o en blancos.</li> </ul>
<b>total</b>	Monto total sin taxes <ul style="list-style-type: none"> <li>• No existen valores nulos o en blanco.</li> <li>• Apto para el análisis del modelo dimensional.</li> </ul>
<b>total_taxes_and_charges</b>	Valor total de taxes y cargos
<b>grand_total</b>	Cantidad final calculada después de impuestos y cargos/descuentos. <ul style="list-style-type: none"> <li>• No existen valores nulos o en blanco.</li> <li>• Apto para el análisis del modelo dimensional.</li> </ul>
<b>discount_amount</b>	Monto de descuento

Tabla 15: Data profiling de la tabla Orden de compra

### Purchase Order Item

Los campos de interes de esta tabla para el análisis correspondiente de este proyecto son:

<ul style="list-style-type: none"> <li>• name(PK)</li> <li>• creation</li> <li>• modified</li> <li>• item_code</li> <li>• item_name</li> <li>• schedulate_date</li> <li>• expected_delivery_date</li> <li>• qty</li> <li>• rate</li> <li>• amount</li> <li>• net_amount</li> <li>• received_qty</li> </ul>	
Campo	Observación
name(PK)	Identificador único, al generarse una nueva orden de compra. <ul style="list-style-type: none"> <li>• No existen valores nulos o en blanco.</li> </ul>
creation	Fecha de creación de la orden de compra.
modified	Fecha de modificación de la orden de compra.
item_code	Codigo del producto o Item. <ul style="list-style-type: none"> <li>• No existen valores nulos o en blanco.</li> </ul>
item_name	Nombre con el cual se registra el producto. <ul style="list-style-type: none"> <li>• No existen valores nulos o en blanco.</li> </ul>
schedule_date	Fecha programada <ul style="list-style-type: none"> <li>• Formato de fecha valido</li> </ul>
expected_delivery_date	Fecha de entrega estimada <ul style="list-style-type: none"> <li>• Formato de fecha valido</li> </ul>
qty	Cantidad de artículos. <ul style="list-style-type: none"> <li>• No existen valores nulos o en blanco.</li> </ul>
rate	Precio del artículo. <ul style="list-style-type: none"> <li>• No existen valores nulos o en blanco.</li> </ul>
amount	Monto según cantidad de artículos en la orden. <ul style="list-style-type: none"> <li>• No existen valores nulos o en blanco.</li> </ul>
net_amount	Precio de venta final del producto. <ul style="list-style-type: none"> <li>• No existen valores nulos o en blanco.</li> </ul>
received_qty	Cantidad recibida <ul style="list-style-type: none"> <li>• Existen valores en blanco.</li> </ul>

Tabla 16: Data profiling de la tabla Articulo de orden de compra

## Especificación de las necesidades analíticas

Utilizando la herramienta ERP next se brindará la solución de las necesidades identificadas en el proceso de compras de la empresa capital general, en el cual el modelo de datawarehouse busca ser un medio de apoyo al momento del análisis de los procesos de compra de la empresa, por lo cual será necesario contar con los reportes de compras donde se darán a conocer los valores totales de los montos y cantidades de compras realizadas por periodos definidos

que se detallan a continuación:

- Cantidad de compras en las cuales se podrá especificar durante que periodo de tiempo estamos obteniendo esa cantidad y por cual proveedor estamos obteniendo ese dato.
- Total del monto de compras por rango de fechas definidas
- Total de productos adquiridos por proveedor
- Total del monto de productos adquiridos por proveedor. visualizando diferentes filtros como descuentos específicos de las compras realizadas con ellos o visualizando por rango de fechas..
- Producto con más y menos compras.

## Modelo dimensional propuesto

### Paso 1 Seleccionar el modelo del negocio

Proceso de compras
--------------------

Tabla 17: Definición de modelo de negocio

### Paso 2 Definir el nivel de granularidad

Granularidad deseada	Se desea ver las compras en función de proveedor, ítems, Datawarehouse por año/mes/trimestre/semestre/año
Granularidad posible	Se pueden ver las compras en función de Items, Proveedor, Datawarehouse por año/mes/trimestre/semestre/año

Tabla 18: Definición de granularidad

### Paso 3 Identificar las dimensiones

Dimensión	Atributos
Item	<ul style="list-style-type: none"> <li>• ItemId(BK)</li> <li>• ItemKey(Sk)</li> <li>• Nombre</li> <li>• ItemGroup</li> </ul>
OrdenCompras	<ul style="list-style-type: none"> <li>• OrderId(bk)</li> <li>• OrderKey(Sk)</li> <li>• CantidadTotal</li> <li>• FechaConfirmaciónOrden</li> <li>• FechaProgramada</li> <li>• Cantidad con impuestos</li> <li>• Cantidad Pagada</li> </ul>
Proveedor	<ul style="list-style-type: none"> <li>• ProveedorId(BK)</li> <li>• ProveedorKey(SK)</li> <li>• Nombre</li> </ul>

Fecha	<ul style="list-style-type: none"> <li>• FechaId(BK)</li> <li>• FechaKey(SK)</li> <li>• Mes</li> <li>• Trimestre</li> <li>• Semestre</li> <li>• Anio</li> <li>• MesNombre</li> </ul>
-------	--

Tabla 19: Definición de dimensiones

#### Paso 4 Identificar Métricas

- Quiero saber cantidad de compras por producto.
- Quiero saber la cantidad de productos comprados por proveedor
- Quiero saber el monto total de compra (Valor puntual de la base transaccional)
- Quiero saber el item que más se compra. (Contabilizar la cantidad de productos y un max)
- Quiero saber el ítem que menos se compra. (Contabilizar la cantidad de productos y un min)
- Quiero saber el monto total de compras por productos. (Este valor se calculará en ejecución dado el dinamismo del DW, dependerá de los parámetros de fechas)
- Quiero saber el monto total de compras por proveedor. (Este valor se calculará en ejecución dado el dinamismo del DW, dependerá de los parámetros de fechas)

Tabla 20: Definición de métricas

### Diagrama de estrella del modelo dimensional

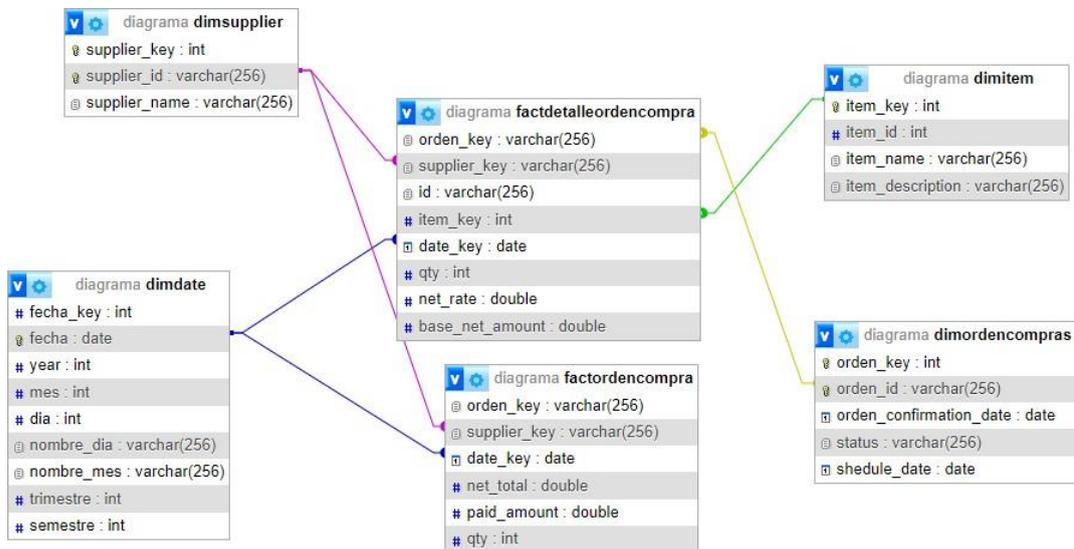


Imagen 9: Diagrama de Estrella Modelo Dimensional

## Mapping de modelo dimensional y tablas del data set

Dim Fecha

Tipo de tabla		Dimensión
<b>Nombre</b>		Dimdate
<b>Nombre visua</b>		Dimdate
<b>Descripción</b>	La DimFecha, contiene el detalle de todas las fechas desde el 2019 al 2025	
<b>Usado en el esquema</b>		

Tabla 21: Dimension Fecha

Objetivo									
Columna	Nombre visual	Descripción	Grupo de atributo	Tipo de dato	tamaño	Precisión	PK	NULL?	Tipo SCD
<b>fecha_key</b>	Fecha key	Llave primaria subrogada	Identificador	bigint			PK	no	0
<b>fecha</b>	fecha ID	Bussines key	Identificador	date				no	0
<b>Nombre_dia</b>	Nombre_dia	Nombre del día	nombre	Varchar	25			no	0
<b>Nombre_mes</b>	Nombre_mes	Nombre del mes	nombre	Varchar	25			no	0
<b>semestre</b>	semestre	Indica el semestre correspondiente	nombre	Int				no	0
<b>year</b>	year	Indica el año correspondiente	nombre	Int				no	0
<b>mes</b>	Mes	Indica el numero en mes	nombre	Int				No	0
<b>trimestre</b>	trimestre	Indica el trimestre correspondiente	nombre	int				no	0

Tabla 22: Mapeo de la Dimension Fecha

DimOrdenCompra

Tipo de Tabla	Dimensión
<b>Nombre</b>	DimOrdenCompra
<b>Visual</b>	dimordencompra
<b>Descripcion</b>	La dimensión DimOrdenCompra , incluye el encabezado de una orden de compra
<b>esquema</b>	CapitalGeneralDW

Tabla 23: Dimension Orden de Compra

Objetivo											Fuente				
Columna	Nombre visual	Descripción	Grupo de atributo	Tipo de dato	Tamaño	Precisión	Llave primaria	NULL ?	Tipo de SCD	Sistema	Esquema	Tabla	Campo	Tipo de datos	Reglas ETL
<b>order_key</b>	Item Key	Llave primaria subrogada	Identificado r	Int			PK	N	0						
<b>order_id</b>	order Id	Bussines key	Identificado r	Varchar	140			N	1	ERP NEXT	CapitalGeneral	tabpurchase order	name	Varchar(140)	
<b>Order_confirmation_date</b>	Order confirmation date	Nombre del articulo	Nombre	Varchar	140			N	1	ERP NEXT	CapitalGeneral	tabpurchase order	order_confirmation_date	date	Convertir a mayuscula
<b>Schedule_date</b>	Shedule_date	Horario de fecha	Nombre	Varchar	140			N	1	ERP NEXT	CapitalGeneral	tabpurchase order	schedule_date	date	Convertir a mayuscula
<b>status</b>	status	estado	Nombre	varchar	20	2		N	1	ERP NEXT	CapitalGeneral	tabpurchase order	status	varchar	

Tabla 24: Mapeo Dimension Orden de Compra

DimSupplier

Tipo de Tabla		Dimension	
Nombre		DimSupplier	
Nombre Visual		dimsupplier	
Descripcion		La dimensión DimSupplier, incluye todos los proveedores de la empresa	
Usado en el esquema		CapitalGeneralDW	

Tabla 25: Dimension Proveedor

Columna	Objetivo										Fuente				
	Nombre visual	Descripción	Grupo de atributo	Tipo de dato	Tamaño	Precisión	Llave primaria	NULL?	Tipo de SCD	Sistema	Esquema	Tabla	Campo	Tipo de datos	Reglas ETL
supplier_key	Supplier Key	Llave primaria subrogada	Identificador	Int			PK	N	0						
supplier_id	Supplier Id	Bussines key	Identificador	Varchar	140			N	1	ERP NEXT	CapitalGeneral	tabPurchase supplier	name	Varchar(140)	
supplier_name	Supplier Name	Nombre del proveedor	Nombre	Varchar	140			N	1	ERP NEXT	CapitalGeneral	tabPurchase supplier	Supplier_name	Varchar(140)	Convertir a mayuscula

Tabla 26: Mapeo Dimension Proveedor

DimItem

Tipo de Tabla		Dimension
Nombre	DimItem	
Nombre Visual	DimItem	
Descripción	La dimensión DimItem, incluye todos los artículos/productos, grupo	
Usado en el esquema	CapitalGeneralDW	

Tabla 27: Dimension Producto

Objetivo											Fuente				
Columna	Nombre visual	Descripción	Grupo de atributo	Tipo de dato	Tamaño	Precisión	Llave primaria	NULL?	Tipo de SCD	Sistema	Esquema	Tabla	Campo	Tipo de datos	Reglas ETL
Item_key	Item Key	Llave primaria subrogada	Identificador	Int			PK	N	0						
Item_id	Item Id	Bussines key	Identificador	Varchar	140			N	1	ERP NEXT	CapitalGeneral	tabItem	name	Varchar(140)	
Item_name	Item Name	Nombre del artículo	Nombre	Varchar	140			N	1	ERP NEXT	CapitalGeneral	tabItem	item_name	Varchar(140)	Convertir a mayuscula

Tabla 28: Mapeo Dimension Producto

## FactOrdenCompras

Tipo de Tabla	Fact-Tabla de hechos
Nombre	FactOrdenCompras
Nombre Visual	factordencompras
Descripción	FactCompras captura transacciones de compra a nivel de la línea de factura
<b>OBJETIVO</b>	

Columna	Nombre Visual	Descripcion	Datatype	Size	Precisi on	Key?	FK To	NULL?	Default Value	Example Values
<b>order_key</b>	Order Key	Llave a dimensión DimOrder	int			FK	DimOrder.DimOrderKey	N		01, 02,33
<b>supplier_key</b>	Supplier Key	Llave a dimensión DimSupplier	int			FK	DimSupplier.DimSupplierKey	N		01,02,03
<b>fecha_key</b>	Fecha Key	Llave a dimensión DimFecha	int			FK	DimFecha.DimFechaKey	N		20221511
<b>qty</b>	qty	Cantidad de compras en una linea	int					N		1,2,3
<b>Paint_moment</b>	Total de la compra unitaria	Total de la compra unitaria	decimal		21,9			N		
<b>Net_total</b>	Total de Compra	Costo total de la compra	decimal		21,9			N		800

Tabla 29: Mapeo tabla de hechos de Orden de Compras

### FactDetalleOrdenCompras

CREATE TABLE public.factdetalleordencompra (

orden\_key character varying(256) ENCODE lzo,  
 supplier\_key character varying(256) ENCODE lzo,  
 id character varying(256) ENCODE lzo,  
 item\_key integer ENCODE az64,  
 date\_key date ENCODE az64,  
 qty integer ENCODE az64,  
 net\_rate real ENCODE raw,  
 base\_net\_amount real ENCODE raw

) DISTSTYLE AUTO;

Tipo de Tabla		Fact-Tabla de hechos								
<b>Nombre</b>	FactDetalleOrdenCompras									
<b>Nombre Visual</b>	factdetalleordencompras									
<b>Descripción</b>	FactDetalleOrdenCompras captura transacciones de compra a nivel de la línea de factura (nivel de detalle)									
<b>OBJETIVO</b>										
Columna	Nombre Visual	Descripción	Datatype	Size	Precision	Key ?	FK To	NULL?	Default Value	Example Values
<b>order_key</b>	Item_key	Llave a dimensión DimOrder	int			FK	DimOrder.DimOrderKey	N		01, 02,33

<b>Item_key</b>	Item_key	Llave a la dimension DimItem	varchar			FK	DimItem	N		
<b>supplier_key</b>	Supplier Key	Llave a dimensión DimSupplier	int			FK	DimSupplier.DimSupplierKey	N		01,02,03
<b>fecha_key</b>	Fecha Key	Llave a dimensión DimFecha	int			FK	DimFecha.DimFechaKey	N		20221511
<b>qty</b>	qty	Cantidad de compras en una línea	int					N		1,2,3
<b>Net_rate</b>	Total de Compra	Costo total de la compra	decimal		21,9			N		
<b>Base_net_among</b>	Total de la compra unitaria	Total de la compra unitaria	decimal		21,9			N		800

Tabla 30: Mapeo de la tabla de hechos de Detalle de orden de compras

### c) Descripción de la tecnología a utilizar

En esta sección se describen las herramientas tecnológicas a utilizar para el desarrollo de la solución propuesta.

- **MariaDB**

Es un sistema de gestión de base de datos relacionales de código abierto con licencia GPLv2, derivado de MySQL. Es uno de los servidores de bases de datos más populares del mundo.



*Imagen 10: Logo del gestor de Base de datos Maria DB*

- **ERPNext**

ERP Next es software ERP open source o código abierto creado por la empresa india Fraptee, un programa instalado en la nube orientado a empresas y que busca dar una solución a la gestión de un modo sencillo y eficaz.

ERPNext está compuesto por módulos como Contabilidad, RRHH, Facturación, Ventas, Compras, CRM, Escritorio o Website.



*Imagen 11: Logo de Software ERPNext*

- **Talend Open Studio**

Talend Open Studio (TOS) es una suite que aporta un conjunto muy complejo, variado y completo de herramientas para llevar a cabo la integración de datos que se ofrece en una versión de código libre (open source). Precisamente por ello, esta es una de las herramientas de integración ETL (extract, transform, load) más utilizadas dentro del mundo Big Data.



*Imagen 12: Logo del software Talend Open Studio*

- **Amazon Web Services (AWS)**

Es la nube más adoptada y completa en el mundo, que ofrece servicios integrales de centros de datos a nivel global.



*Imagen 13: Logo de Amazon Web Services*

- **Amazon Redshift**

Amazon Redshift es un servicio de data warehouse administrado de varios petabytes en la nube.

Un almacenamiento de datos de Amazon Redshift es un sistema de administración y consulta de bases de datos relacionales de clase empresarial. Amazon Redshift admite las conexiones de clientes con muchos tipos de aplicaciones, incluidas las herramientas de análisis, datos, generación de informes e inteligencia empresarial (BI).



*Imagen 14: Logo de Amazon Redshift*

- **Amazon S3**

Amazon S3 es un servicio de almacenamiento de objetos que ofrece escalabilidad, disponibilidad de datos, seguridad y rendimiento.



*Imagen 15: Logo de Amazon S3*

- **Amazon IAM**

Es un servicio web que lo ayuda a controlar de forma segura el acceso a los recursos de AWS. administrar de forma centralizada los permisos que controlan a qué recursos de AWS pueden acceder los usuarios.



*Imagen 16: Logo de IAM*

- **Power BI**

Es un servicio gratuito de análisis de negocio basado en la nube y visualización de datos, de negocio. Esta herramienta de Business Intelligence (BI), incorporada en la suite de productividad Microsoft Office 365, permite crear dashboard, informes o reportes interactivos con Power BI Desktop y acceder a los datos en cualquier lugar con las aplicaciones nativas en su versión móvil.



*Imagen 17: Logo de Power BI*

## d) Diagrama Arquitectónico de la solución

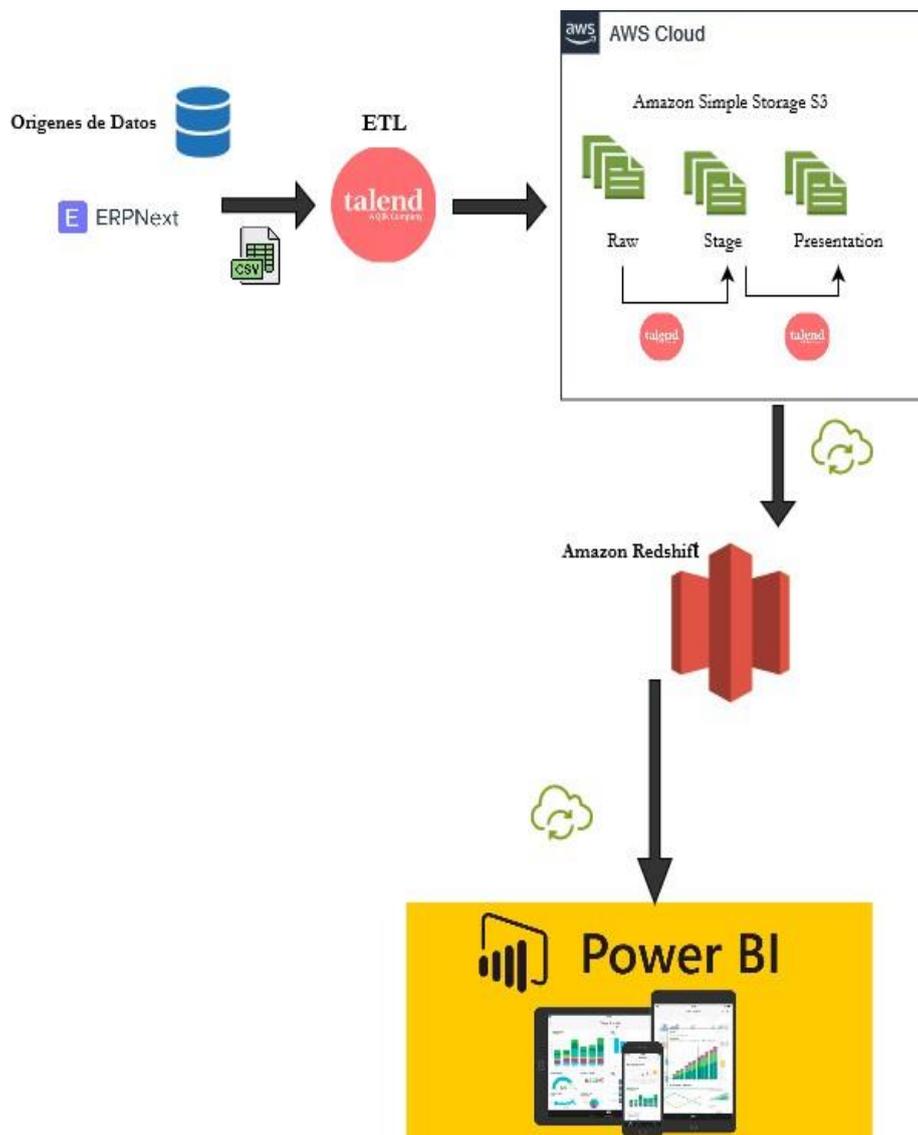


Imagen 18: Diagrama arquitectónico de la solución

## e) Descripción de cada componente de la solución

### Orígenes de Datos

Para obtener los datos para un Datawarehouse existen distintas fuentes tales como archivos, bases de datos, sistemas transaccionales, aplicaciones web, sistemas operativos entre otras. La empresa Capital General ha utilizado los siguientes componentes:

- ERP NEXT: La empresa Capital general utiliza este sistema para automatizar sus procesos empresariales, así como registrar un proveedor, ingresar las órdenes de compra. Obteniendo toda la información almacenada sobre el proceso de compras que tiene la empresa.
- CSV: ERP Next genera los datos en archivos csv, los cuales luego se utilizaran para los procesos ETL.

## **ETL**

Es necesario para la carga de datos a un Datawarehouse, se necesitan distintas tareas de extracción transformación y carga (ETL), como verificación de tipos de datos, formatos para mejorar la calidad y la integridad de los mismos. La herramienta que se utilizó para la solución fue la siguiente:

- Open Talend: Aporta una completa ayuda para la integración de los datos, con extrayendo los datos del sistema origen, luego se limpian los datos para mejorarlos y luego se pasa a la aparte de carga de datos a las dimensiones para llegar a la etapa final de la presentación. Almacenándose todo a través de Amazon S3

## **AWS**

Siendo un servicio en la nube completa en la que ofrece soluciones para las bases de datos, almacenamiento, y análisis de los datos. Nos proporciona ventajas tales como integración con herramientas de análisis, costos bajos, accesibilidad a los datos, estabilidad y fiabilidad, por lo cual para la ejecución del proyecto se usaron los siguientes componentes:

Amazon S3: Herramienta para almacenamiento y recuperación de volúmenes de datos. Durante el desarrollo del proyecto se crearon las 3 zonas para almacenar y procesar la información, las cuales son:

- RAW: Almacenó los datos tal y como vienen del origen de datos.
- STAGE: Procesó y almaceno los datos para llevarse transformados al área de PRESENTACION, dichas tareas fueron como verificar duplicados y cálculos necesarios
- PRESENTACION: Se almacenaron los datos ya procesados para poder presentarse para el respectivo análisis y futuras tomas de decisiones.

La información almacenada en la etapa de presentación se carga a Amazon Redshift a través de procesos ETL y la herramienta OpenTalend para facilitar el análisis de los datos.

## **AMAZON REDSHIFT**

Puesto que es un servicio de Datawarehouse, crearemos nuestro modelo dimensional, almacenando los datos de la zona de presentación de S3 ayudando a un rendimiento efectivo y facilitando la conexión a power BI.

## **AMAZON IAM**

Se implemento para la gestión de usuarios, roles y permisos que fueron asignados a los usuarios correspondientes.

## **Visualización**

Cuando los datos ya han sido procesado y almacenados de esta forma, se vuelve necesario representarlos de manera visual para un mejor análisis. Por lo que se empleo el uso de la herramienta Power BI

Power BI

Utilizada para la representación gráfica de los datos, ya que permite crear dashboard compuestos de gráficos, informes, tablas para mayor facilidad de entendimiento de la información ya antes transformada.

Por lo que por medio de estos reportes se solucionan los requerimientos planteados por la empresa capital general, presentando visualmente las distintas métricas que ayudaran a futuros análisis, identificación de problemas y toma de decisiones para un mejor manejo de los procesos de compras.

## Capítulo III: Estrategia de Implementación de Propuesta de Solución

### a) Estrategia de Implementación

Esta sección proporciona una descripción concisa de los elementos que participan en el almacenamiento, organización, extracción, transformación y carga de los datos transaccionales de la compañía en el Data Warehouse. Además, se presentan los resultados obtenidos de la implementación de esta solución, mostrando capturas de pantalla del aplicativo. Estas capturas están clasificadas según la tecnología empleada y su función en las distintas áreas del proceso de tratamiento de datos.

### Base de datos Mysql

El gestor de bases de datos Mysql tiene una función en el procesamiento de datos, alojar los datos transaccionales de la orden de compras de la empresa.

**Base de datos transaccional** Fuente de datos transaccionales a los que se le aplicaran los procesos de extracción, transformación y carga. En nuestro caso hemos mapeado la base de datos para poder obtener la estructura y realizar el analisis correspondiente.

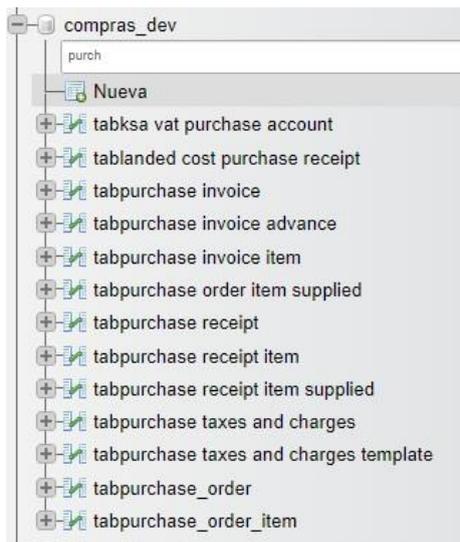


Imagen 19: Base de datos transaccional

### Bucket S3

El bucket de S3 es el componente principal de almacenamiento y organización de archivos, hace la función de Data Lake y contiene los archivos necesarios para poder procesar los archivos.

**Procesados:** Contiene los archivos que ya procesados por los ETL, se utiliza para llevar un histórico de las cargas.

**Carpeta temporal:** Es una carpeta que nos sirve para agregar diferentes archivos que se consideren necesarios, pero de uso temporal.

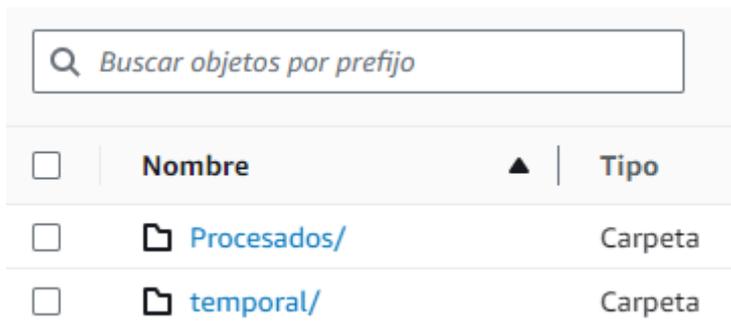


Imagen 20: Estructura de carpetas en Amazon S3

## Redshift

Componente principal de almacenamiento de los datos del Data Warehouse, la base de datos está construida utilizando un esquema de estrella que consta de 4 dimensiones y dos tablas de hechos, este modelo de datos estructurados le permite aprovechar el procesamiento de datos y facilita la lectura eficiente de millones de registros de las órdenes de compra, además, permite el consumo de la información y la visualización su comportamiento en Power BI

orden_key	orden_id	orden_confirmation_date	status	shedule_da
1	PUR-ORD-2023-00973	2023-02-16	Completed	2023-03-02
2	PUR-ORD-2023-00972	2023-02-14	To Receive and Bill	2023-02-28
3	PUR-ORD-2023-00971	2023-02-13	To Receive and Bill	2023-02-27
4	PUR-ORD-2023-00970	2023-02-10	To Receive and Bill	2023-02-24
5	PUR-ORD-2023-00969	2023-02-08	To Receive and Bill	2023-02-22
6	PUR-ORD-2023-00968	2023-02-03	To Receive and Bill	2023-02-17
7	PUR-ORD-2023-00967	2023-01-30	To Receive and Bill	2023-02-13

Imagen 21: Base de datos en Amazon Redshift

## Proceso ETL en Talend

A continuación, se describirán los trabajos más importantes que se encargan de la transformación de los datos.

**Aclaraciones:** Cada imagen agregada representa un flujo de los jobs realizados, se muestra la estructura como tal.

## Ingresar dimfecha

La dimfecha es un csv que el equipo armo y se basa en parámetros de años 2019 al 2025, y inicialmente esta subida en la carpeta raw, y el job se encarga de pasar el archivo a la carpeta stage para el procesamiento y subida a redshift, además registra un log en una base de datos aparte llamada "logs" para llevar el registro de los movimientos.

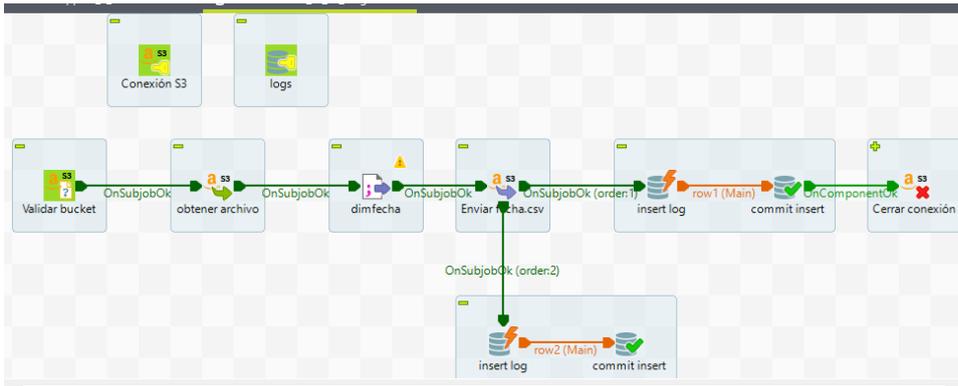


Imagen 22: Proceso ETL – Carga de datos a carpeta Raw y Stage

### Log generado

Se ha subido la dimfecha a la carpeta stage	2023-12-08 19:34:31
Se ha subido la dimfecha a la carpeta stage	2023-12-08 19:40:15

Imagen 23: Log generado al ejecutar el proceso ETL

### Dimfecha hacia redshift

El job se encarga de obtener la el csv de la dimdate y mapearlo en redshift , si llegara a existir inconvenientes a la hora de mapear, se generar un log con el error, y si no se registra el log de éxito.

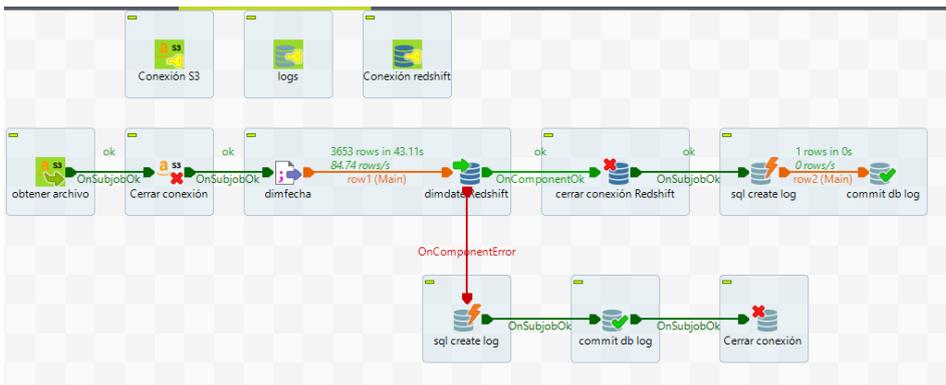


Imagen 24: Proceso ETL - Carga de datos en Redshift

### Archivo de Supplier.csv de la carpeta raw a la stage

Se decidió que los catálogos se migren por completo entonces el job se encarga de enviar el archivo de supplier.csv de la carpeta raw a la carpeta stage pero ya renombrado para poder cargarlo a redshift. Y se generan sus respectivos logs

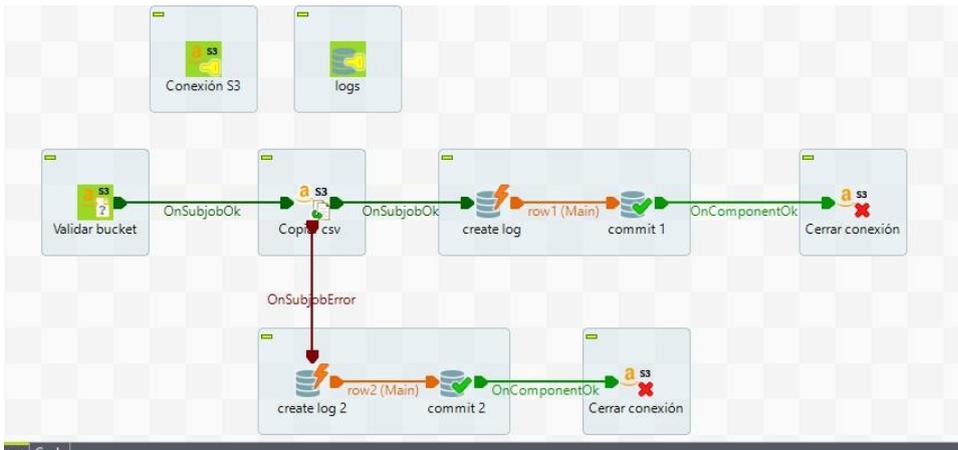


Imagen 25: Proceso ETL - Carga de datos en la carpeta Raw y Staging

### Cargar dimsupplier a redshift

Job que se encarga de subir el archivo de catalogo de proveedores de la carpeta raw a stage, preparandolo para poder procesarlo, de igual manera tiene sus respectivos puntos de log.

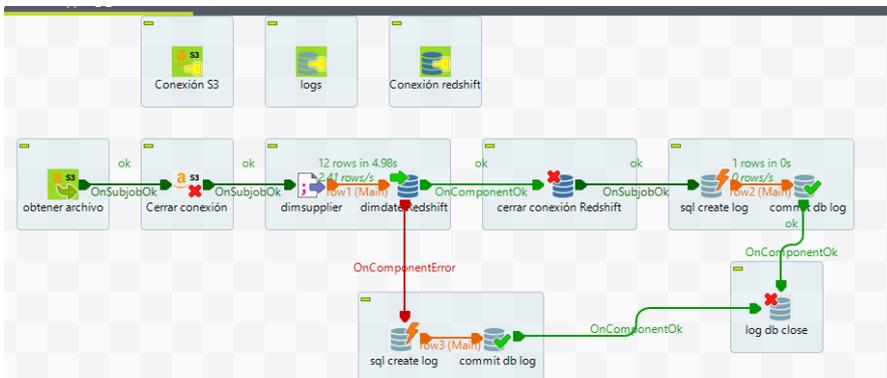


Imagen 26: Proceso ETL - Carga de datos en Redshift

### Archivo de Item.csv de la carpeta raw a la stage

Job se encarga de enviar el archivo de item.csv de la carpeta raw a la carpeta stage pero ya renombrado para poder cargarlo a redshift. Y se generan sus respectivos logs

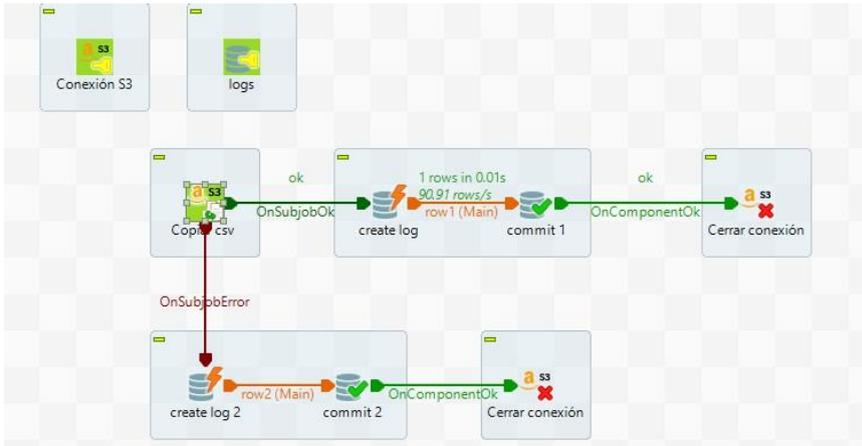


Imagen 27: Proceso ETL - Carga de datos a la carpeta Raw y Stage

### Cargar dimitem a redshift

Job que se encarga de subir el archivo de catalogo de productos de la carpeta raw a stage, preparándolo para poder procesarlo, de igual manera tiene sus respectivos puntos de log.

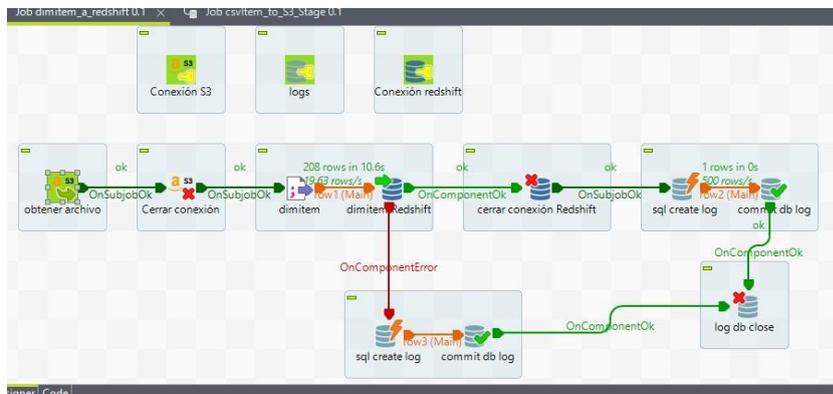


Imagen 28: Proceso ETL - Carga de datos a Redshift

### Archivo de PurchaseOrden.csv de la carpeta raw a la stage

Job se encarga de enviar el archivo de PurchaseOrden.csv de la carpeta raw a la carpeta stage pero ya renombrado para poder cargarlo a redshift. Y se generan sus respectivos logs

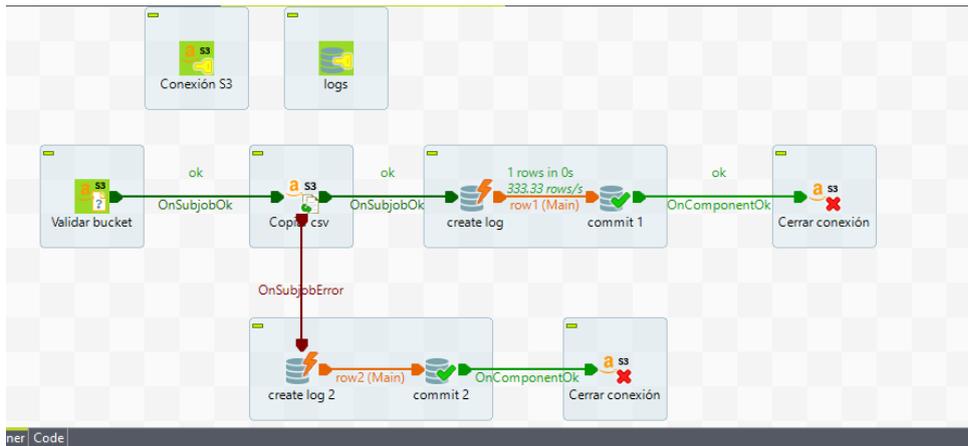


Imagen 29: Proceso ETL - Carga de datos a la carpeta Raw y Stage

### Cargar dimordencompra a redshift

Job que se encarga de subir el archivo que tiene las ordenes de compras (emcabeazado), preparándolo para poder procesarlo, de igual manera tiene sus respectivos puntos de log. Por ser uno de los archivos de mayor masa, se dispusieron varios logs para controlar los errores, además de un tmap para mapear la información.

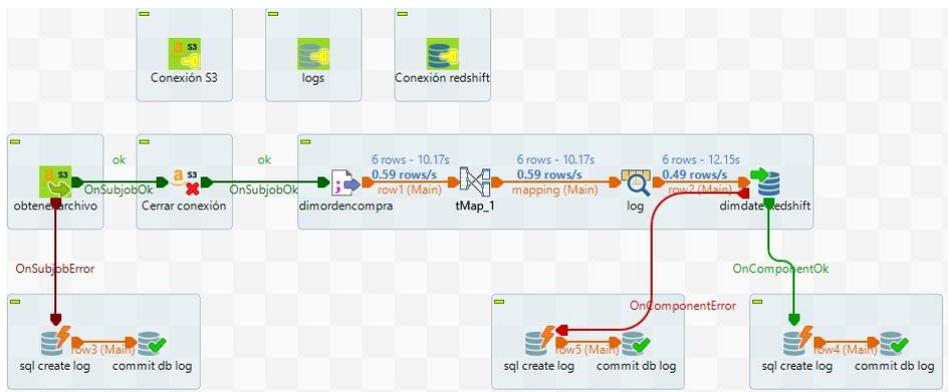


Imagen 30: Proceso ETL -Carga de datos a Redshift Dimordencompra

### Archivo de PurchaseOrderFact.csv de la carpeta raw a la stage

Job se encarga de enviar el archivo de **PurchaseOrderFact.csv** de la carpeta raw a la carpeta stage pero ya renombrado para poder cargarlo a redshift. Y se generan sus respectivos logs

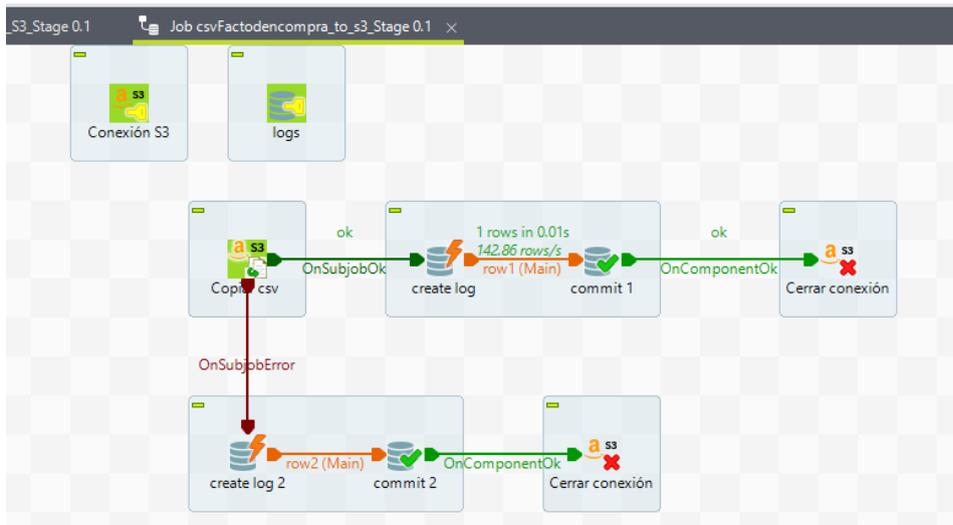


Imagen 31: Proceso ETL - Carga de datos a la carpeta Raw y Stage

Un detalle a comentar es que los datos por facilidad ya se convierten desde el propio ERP , sin embargo, en un ítem posterior se explicara a mayor detalle el funcionamiento y como se generan los archivos csv para poder procesar la data.

### Cargar factordencompra a redshift

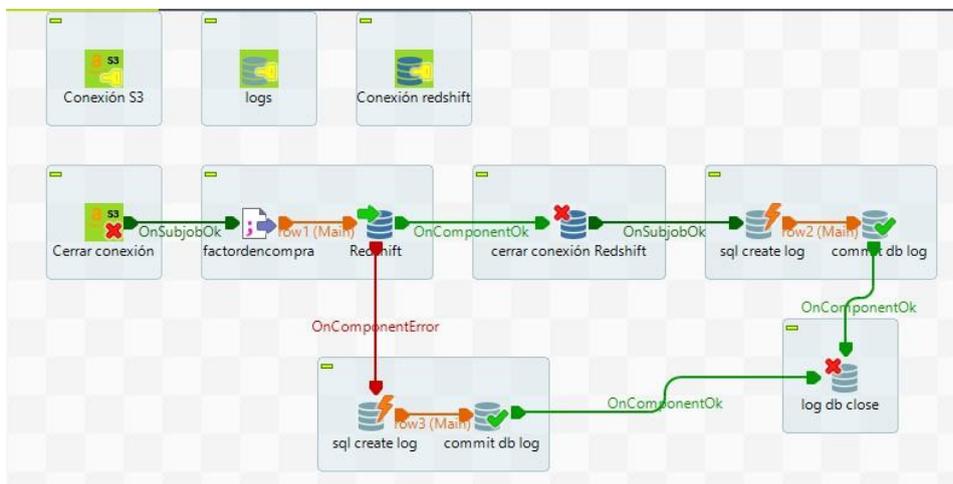


Imagen 32: Proceso ETL - Carga la Fact orden de compra a Redshift

### Procedimiento alternativo para correr dimordencompra y factordencompra

Este job se encarga de tomar un csv que está en la carpeta raw, que es el que representa las órdenes de compra (cabecera) pero a diferencia de los job anteriores para cargar los archivos dimordencompra y factordencompra a la carpeta stage, aquí lo hace a partir de un archivo completo generado desde el erp con los datos de las órdenes. El map se encarga de dividir el archivo y crear dos bloques para poder cargar los dos archivos a la carpeta stage. Al finalizar ya tendremos dos archivos listos para cargar a redshift sin necesidad de dividirlo manualmente. Ya partiendo de aquí se ejecuta los jobs para cargar la dim y la fact de forma normal para cargar la data en redshift

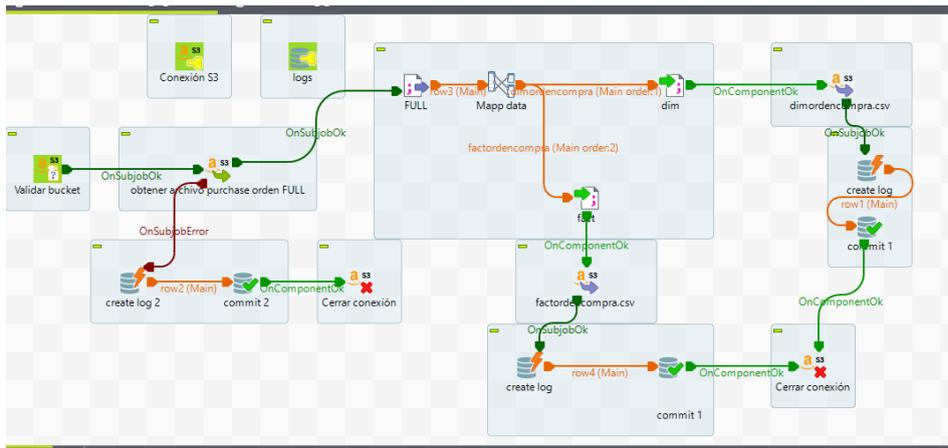


Imagen 33: Proceso ETL alternativo para la carga de datos de dimordencompra y factordencompra

Como se mencionaba anteriormente se descarga la data de las ordenes de forma general en el csv y ya el job alternativo se encarga de hacer la tarea de dividir la información correspondiente

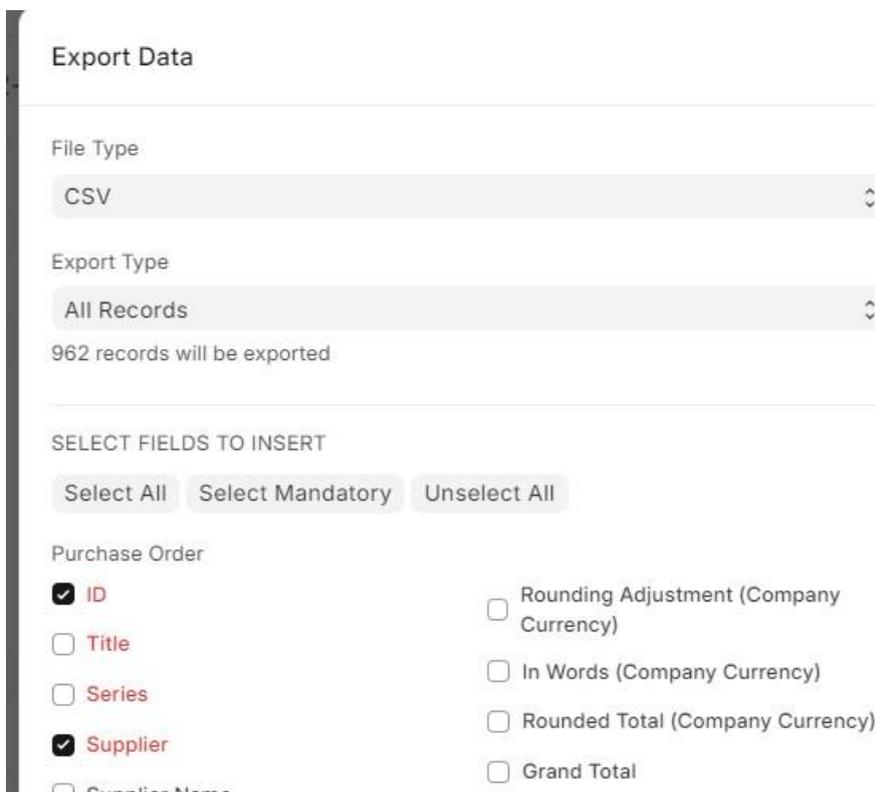


Imagen 34: Exportación de datos desde ERPNext

### Procedimiento para cargar factdetalleordencompra

Job que se encarga de mover y transformar la información de la carpeta raw a la stage que es la que se utilizara para subir a redshift. Pero en este caso hay que transformar un poco la información. Se descarga el archivo desde S3 y luego lo almacenamos en un csv para poder manipular la data, la transformamos en el Tmap y luego se sube el nuevo csv a s3.

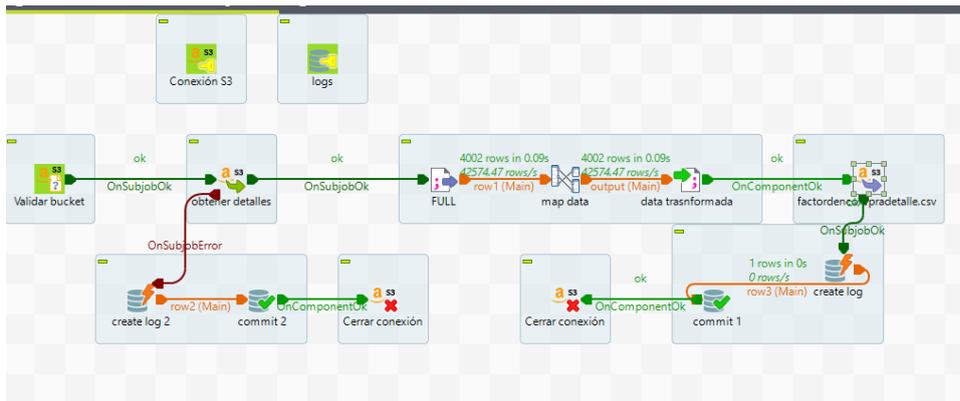


Imagen 35: Proceso ETL - Carga de datos en la carpeta Raw y Stage de factdetalleordencompra

Ejemplo del csv raw generado desde el erp, como se visualiza se debe completar en el csv los campos vacíos de la orden\_key y supplier\_key, ya que así de esta forma será muy complicado pasarlo a redshift , ya que, si bien al leer el csv nosotros le entendemos, a la hora de procesarlo es más complicado.

	A	B	C	D	E	F	G
1	orden_key	supplier_key	id	item_key	qty	net rate	base net amount
2	PUR-ORD-2023-00973	INTEGRAT00	7934a440c3	18	12	7248.48	604.04
3			e2fb2c2c6f	125	10	5940.4	594.04
4			18645780a2	21	15	9148.35	609.89
5			f48c674048	128	10	6504.8	650.48
6			5871baa790	14	10	6336.5	633.65
7	PUR-ORD-2023-00972	ELECTRON0	0266a090de	37	15	16029.9	1068.66
8			edf447c27c	38	10	10195.1	1019.51
9			c680dc29ba	195	9	7524.27	836.03
10	PUR-ORD-2023-00971	ELECTRON0	b524980ae4	265	12	13961.88	1163.49
11			c2a1a56da0	279	10	11398.7	1139.87
12			b29fc6e55d	36	15	13694.7	912.98
13	PUR-ORD-2023-00970	ILLINOIS00	458cd3824f	147	15	3609.3	240.62
14			d30e86ad54	149	15	3215.4	214.36
15			eabbd3f568	150	15	3358.8	223.92
16			fbcb68c746a	261	10	13302.6	1330.26
17			4a98636ec7	262	10	10519.7	1051.97
18	PUR-ORD-2023-00969	ANDERSON0	0a14664e2c	41	10	2315.8	231.58
19			b6d3d29783	43	15	3992.25	266.15
20			9a354d009b	134	10	2523.1	252.31

Imagen 36: Archivo csv generado desde ERPNext

El resultado debe ser así

A	B	C	D	E	F	G	H
orden_key	supplier_key	item_key	id	date_key	base_net_amount	qty	net_rate
PUR-ORD-2023-00973	INTEGRAT00	18	7934a440c3	2023-03-02	7248.48	12	604.04
PUR-ORD-2023-00973	INTEGRAT00	125	e2fb2c2c6f	2023-03-02	5940.4	10	594.04
PUR-ORD-2023-00973	INTEGRAT00	21	18645780a2	2023-03-02	9148.35	15	609.89
PUR-ORD-2023-00973	INTEGRAT00	128	f48c674048	2023-03-02	6504.8	10	650.48
PUR-ORD-2023-00973	INTEGRAT00	14	5871baa790	2023-03-02	6336.5	10	633.65
PUR-ORD-2023-00972	ELECTRONO	37	0266a090de	2023-02-28	16029.9	15	1068.66
PUR-ORD-2023-00972	ELECTRONO	38	edf447c27c	2023-02-28	10195.1	10	1019.51
PUR-ORD-2023-00972	ELECTRONO	195	c680dc29ba	2023-02-28	7524.27	9	836.03
PUR-ORD-2023-00971	ELECTRONO	265	b524980ae4	2023-02-27	13961.88	12	1163.49
PUR-ORD-2023-00971	ELECTRONO	279	c2a1a56da0	2023-02-27	11398.7	10	1139.87
PUR-ORD-2023-00971	ELECTRONO	36	b29fc6e55d	2023-02-27	13694.7	15	912.98
PUR-ORD-2023-00970	ILLINOIS000	147	458cd3824f	2023-02-24	3609.3	15	240.62
PUR-ORD-2023-00970	ILLINOIS000	149	d30e86ad54	2023-02-24	3215.4	15	214.36
PUR-ORD-2023-00970	ILLINOIS000	150	eabbd3f568	2023-02-24	3358.8	15	223.92
PUR-ORD-2023-00970	ILLINOIS000	261	fbcb68c746a	2023-02-24	13302.6	10	1330.26
PUR-ORD-2023-00970	ILLINOIS000	262	4a98636ec7	2023-02-24	10519.7	10	1051.97
PUR-ORD-2023-00969	ANDERSONO	41	0a14664e2c	2023-02-22	2315.8	10	231.58
PUR-ORD-2023-00969	ANDERSONO	43	b6d3d29783	2023-02-22	3992.25	15	266.15
PUR-ORD-2023-00969	ANDERSONO	134	9a354d009b	2023-02-22	2523.1	10	252.31
PUR-ORD-2023-00969	ANDERSONO	140	48a58a5cf1	2023-02-22	3599.25	15	239.95
PUR-ORD-2023-00969	ANDERSONO	144	b0c8df637f	2023-02-22	3295.35	15	219.69
PUR-ORD-2023-00969	ANDERSONO	145	b5d6aaaf2f9	2023-02-22	3871.5	15	258.1
PUR-ORD-2023-00968	PROSF0001	247	ad140fb92d	2023-02-17	2893.8	12	241.15

Imagen 37: Archivo csv completado

### Transformación desde map

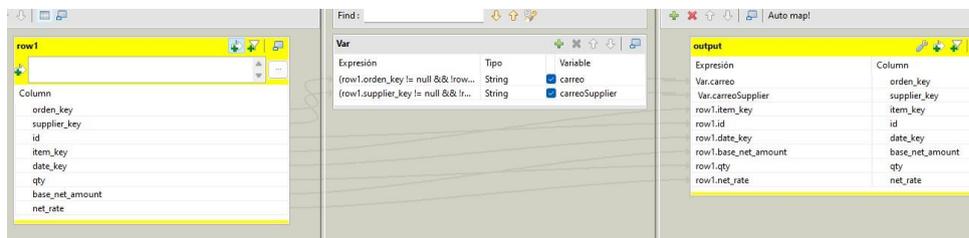
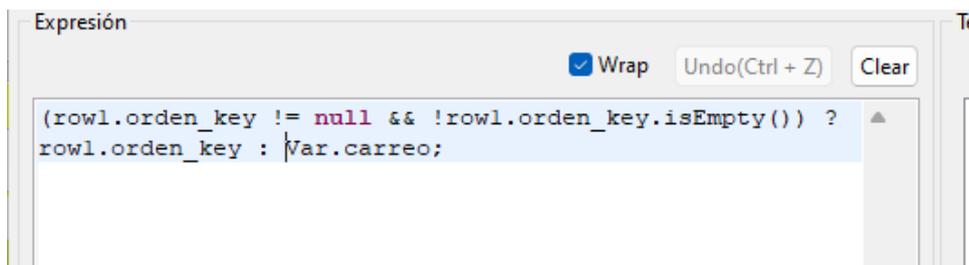


Imagen 38: Transformación de datos usando el componente de mapeo de Talend Open Studio

Expresión lógica utilizada, esta misma aplica para el de supplier, solo sustituyendo los campos convenientes.



### Cargar factdetalleordencompra a redshift

Job se encarga de distribuir el archivo de las ordenes de compras (detalles) hacia redshift, ya que en el anterior job se mapeo y se hizo la distribución, en este job solo se encarga de pasar el csv a redshift.

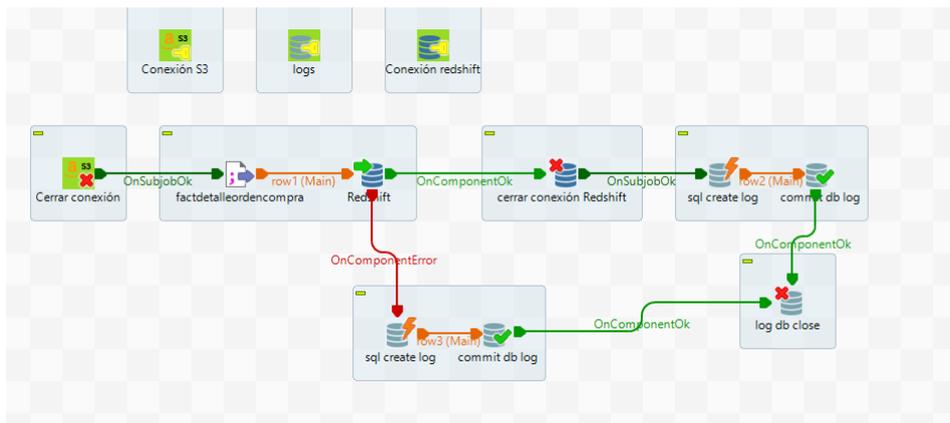


Imagen 39: Proceso ETL – Carga de la factdetalleordencompra a Redshift

### Datos importantes a tener en cuenta sobre la base de datos relacional

La base de datos relacional se obtuvo nada más para poder realizar y análisis de interpretación de los datos para poder realizar el datawarehouse. La información se extrae vía un módulo propio del ERP que permite extraer la información en archivos csv para luego hacer el volcado de datos. Dichos csv son los que se suben a la carpeta raw en Amazon S3, el único cambio que sufren son los headers que se les modifica para mayor facilidad a la hora de hacer el ETL.

Hay tres formas de tener los servicios del ERP next, la primera es descargando una OVA y montarlo en un servidor propio, esto involucra realizar la serie de configuraciones correspondientes de la OVA (que muchas veces ya viene con todos los elementos listos para funcionar pero algunas veces no, dependiendo de donde se obtenga, ya que dichas OVA son generadas por los mismos usuarios que las comparten para que otros usuarios no tengan la dificultad de instalarlo desde cero, sea instalando el entorno, base de datos y otros).

La segunda opción es montarlo desde el proyecto base es decir del repositorio oficial, compilar y montarlo en un servidor Linux con todas sus configuraciones correspondientes.

Y la última es adquirir una membresía mensual en el sitio oficial de ERP next, ellos ofrecen ya el ERP instalado para utilizar.

De las tres opciones las primeras dos sí se puede tener acceso a la base de datos, pero la tercera no, por no ser un servidor propio, y solo se tiene acceso directo al ERP como tal a la interfaz y no al servidor, sin embargo, se busca una solución que involucre al 100% de usuarios, es decir que la solución sea factible para la mayoría, y se obtiene por obtener la información directa de la base de datos, los usuarios que adquieren los planes no podrían emplear dicha solución.

Se analizó la forma en que el ERP generaba los archivos csv con la data y nos pareció una forma muy ordenada y facilitaba el trato de los archivos, los cuales por tales razones se facilitó el poder mapear y procesar la data en Talend.

De tal forma, la solución trabaja a base de los archivos csv que genera el ERP.

## Base de datos de logs

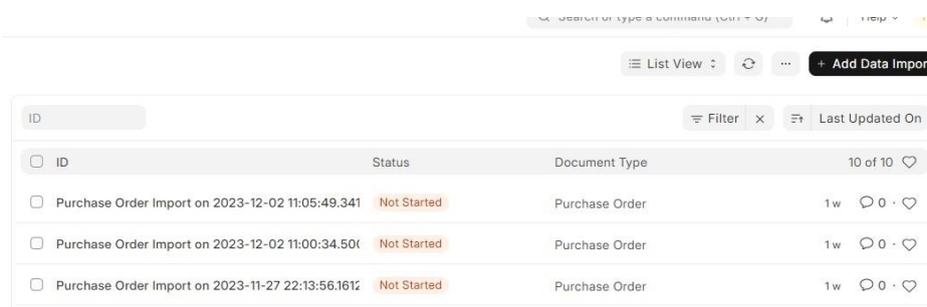
Es una base de datos montada en mysql que tiene una sola tabla, y registra todos los movimientos en los jobs, sea de registro de éxito o de error, así se tiene un mayor control.

```
CREATE TABLE `logs_s3` (  
  `id` int NOT NULL,  
  `comentario` text COLLATE utf8mb4_bin NOT NULL,  
  `fecha` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP,  
  `tipo` varchar(100) COLLATE utf8mb4_bin DEFAULT NULL  
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_bin;
```

Imagen 40: Tabla creada para el registro de logs

## ¿Como se obtiene la data desde el erp?

Para poder acceder debemos entrar con nuestro usuario y dirigirnos en el menu lateral “Tools” y luego dar clic en “import data”, y tendremos que ver una información similar a esta, donde lista todos los procesos de importación que se han tenido.



The screenshot shows the 'Import Data' interface in ERPNext. At the top, there is a search bar and a '+ Add Data Import' button. Below this is a table with columns for ID, Status, Document Type, and Last Updated On. The table contains three rows of import jobs, all with a status of 'Not Started' and a document type of 'Purchase Order'.

ID	Status	Document Type	Last Updated On
Purchase Order Import on 2023-12-02 11:05:49.341	Not Started	Purchase Order	1w
Purchase Order Import on 2023-12-02 11:00:34.500	Not Started	Purchase Order	1w
Purchase Order Import on 2023-11-27 22:13:56.161	Not Started	Purchase Order	1w

Imagen 41: Visualización desde el panel de ERPNext

Luego se le tiene que dar clic en Add data import, y deberá aparecer la opción para crear el registro de importación, en el campo Document Type se debe poner el tipo de elemento a exportar, en este caso es Item (Productos) y en import Type debe de ser insert new records.



The screenshot shows the form for adding data to import. It has two main sections: 'Document Type' and 'Import Type'. The 'Document Type' field is set to 'Item' and has a 'Don't Send Emails' checkbox checked. The 'Import Type' field is set to 'Insert New Records'.

Document Type \*  Don't Send Emails

Item

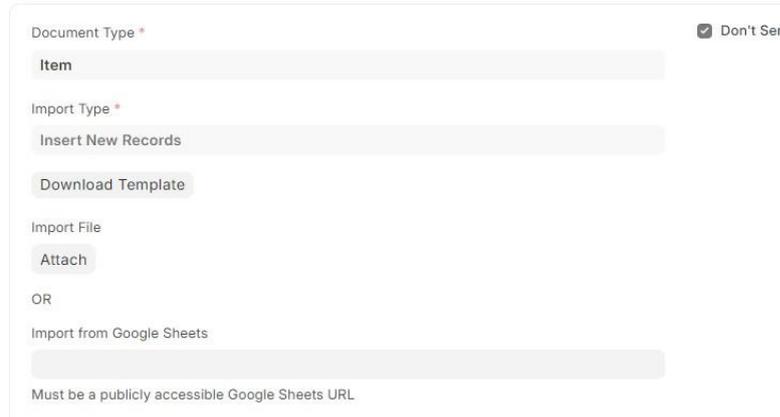
Import Type \*

Insert New Records

Imagen 42: Visualización para agregar la data a importar desde ERPNext

Luego de ello deber aparecer un formulario como este, en él se debe seleccionar el boton de Download Template y es donde les permitirá seleccionar que campos deben exportar y de que tabla hacerlo, tengan en cuenta que pueden seleccionar importación de una o más tablas relacionadas.

Item Import on 2023-11-27 22:08:18.092688 Not Started



Document Type \*  Don't Ser

Item

Import Type \*

Insert New Records

Download Template

Import File

Attach

OR

Import from Google Sheets

Must be a publicly accessible Google Sheets URL

Imagen 43: Importacion por tablas o campos

## Importaciones

### Productos

Document Type: Item

Campos a seleccionar: Item Code, Item Name, Description

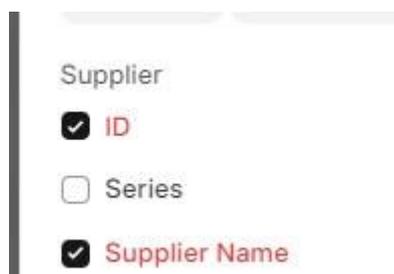
- Item
- Series
  - Item Code
  - Item Name
  - Item Group

Imagen 44: Campos de productos a importar

### Proveedores

Document Type: Supplier

Campos a seleccionar: ID, Supplier Name



Supplier

- ID
- Series
- Supplier Name

Imagen 45: Campos de Proveedores a importar

## Orden de compras (Cabecera)

Document Type: Purchase Orden

Campos a seleccionar: ID, Supplier, Date, RequiredBy, Total Quantity, Net Total, Total, status

## Orden de compras (detalle)

Document Type: Purchase Orden

Campos a seleccionar en la tabla Purchase Orden: ID, Supplier

Campos a seleccionar en la tabla Purchase Orden Item: ID, Item Code, Quantity, Amount, Rate

## Power BI

Se implementará la creación de informes mediante la herramienta power Bi ya que permite utilizar diferentes fuentes de datos y unirlos en una sola, para la realización de estos reportes que ayudaran al análisis de la información transformada.

Se deberá hacer la conexión a la base de datos que se tiene en AWS Redshift para poder acceder a nuestro datawarehouse de la solución.

A continuación, se muestran los pasos a seguir:

- Obtenemos los datos de AWS Redshift, como se muestra en la imagen damos clic en obtener datos selección Redshift y conectamos:

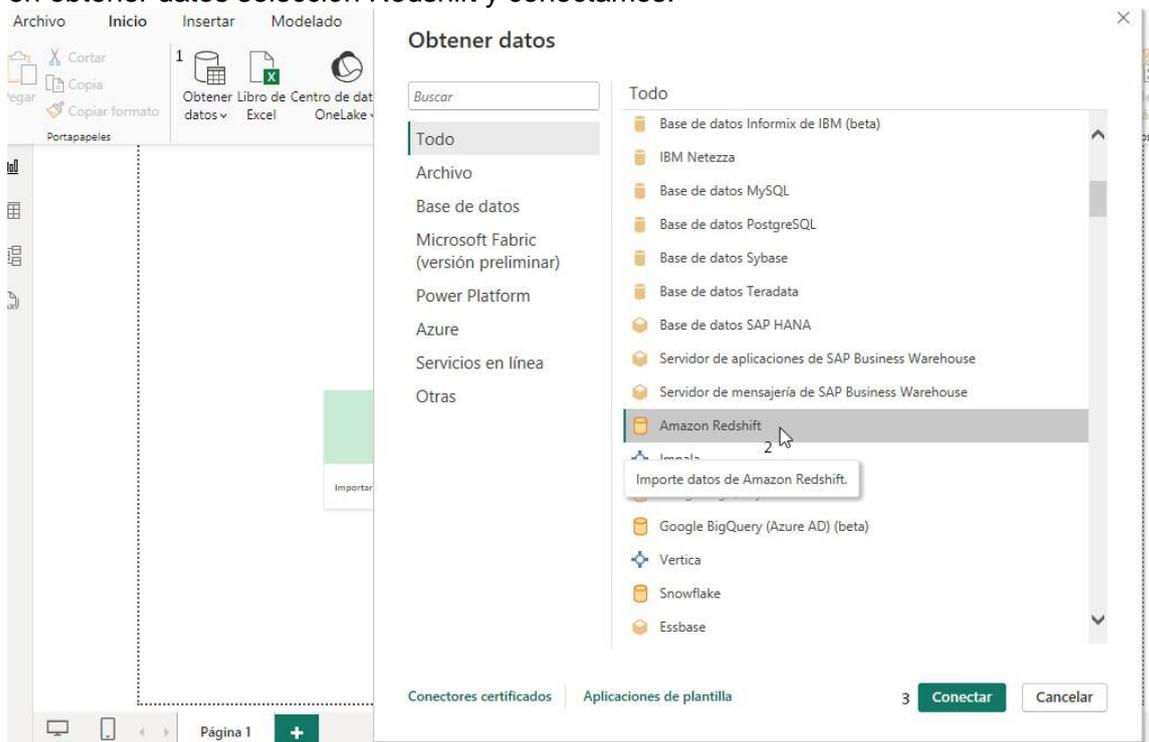


Imagen 46: Conexión a Amazon Redshift desde Power BI

- Luego ingresamos la ruta del servidor configurado en AWS y la base de datos que deseamos conectarnos:



Imagen 47: Configuración del servidor y base de datos de aws Redshift

- Seleccionamos las tablas de nuestro DataWarehouse que utilizaremos para nuestros informes y seleccionaremos cargar:

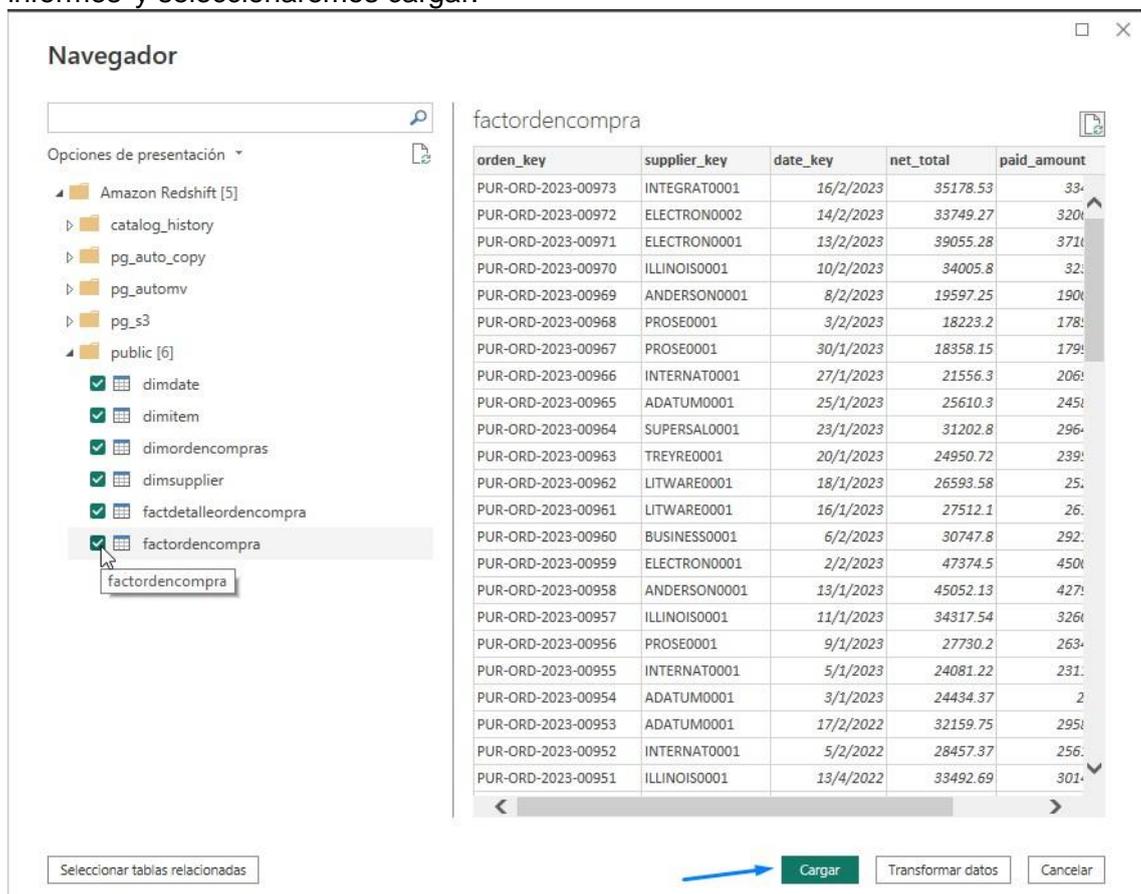


Imagen 48: Selección de tablas a cargar en Power BI

Por lo que, a partir de acá mediante la representación gráfica de las métricas y los resultados obtenidos por ellas, se podrá tener de una mejor manera y a la vez intuitiva esta información para el análisis de los datos que conlleva a una toma de decisiones de parte de la empresa.

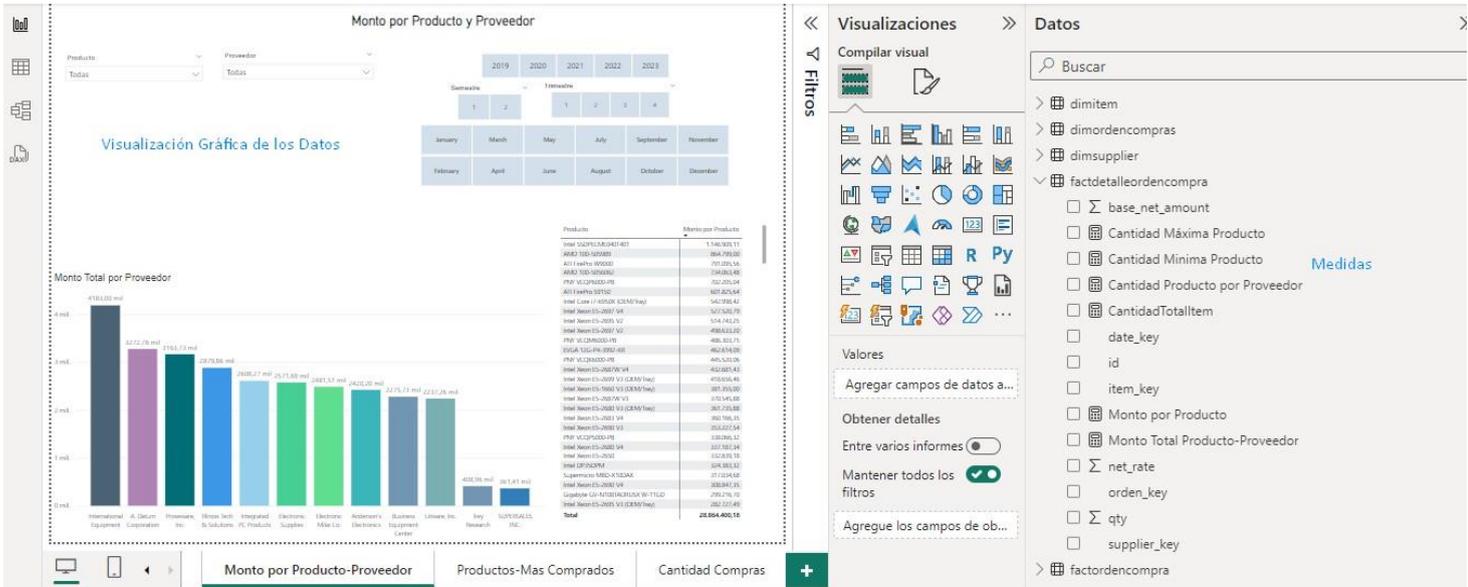


Imagen 49: Visualización de un reporte en Power BI

### Plan de migración

Para el orden de ejecución que se debe seguir es necesario aclarar que primero deben ser los datos tales como producto y proveedor, a partir de ellos continuar con las ordenes de compra que se ejecutara.

Para la dimensión se fecha se general a partir de un archivo de Excel, csv con el cual tiene los componentes necesarios para filtrar los datos a partir de ellos. Puesto que se necesita tener cierto cuidado con esta dimensión para evitar tener problemas a futuro a la hora del análisis de los datos.

Debe seguirse el siguiente orden a la hora de ejecutar los Jobs:

1. Obtener los datos de ERP NEXT en archivos CSV hacia la zona raw
2. Luego cargar los datos de la zona raw a la zona stage
3. Y de la zona stage a la zona de presentación
4. A partir de esta fase cargar hacia Amazon Redshift
5. Y luego cargar a nuestra herramienta de visualización que en este caso será Power BI.

### Frecuencia

Dentro de la solución se ha definido actualizar y descargar los nuevos archivos cada mes por motivo del proceso de compras, y realizar las respectivas actualizaciones en el periodo de menos actividad en la empresa para evitar contratiempos en el sistema.

### b) Presupuesto de Implementación (Distribuido en rubros)

Para el presupuesto de implementación se han considerado únicamente recursos tecnológicos de licenciamiento y servicios en la nube de Amazon.

## Licenciamiento Microsoft Power BI Pro

Para la implementación y puesta en marcha de esta solución se considera adquirir el licenciamiento de Power BI Pro, la cual permitirá crear, publicar, compartir informes con gráficos, y paneles modernos de los análisis de la organización. Actualmente la licencia tiene un valor de \$10.13 por usuario por mes.

Si consideramos que la empresa requiere de dos usuarios analista de datos para la creación y el mantenimiento de dashboards y reportes, es necesario adquirir una licencia de Power BI Pro para cada usuario. Por tanto, el presupuesto de licenciamiento se presenta en la siguiente tabla.

Licencia	Cantidad	Precio	Subtotal
Licencia Power BI Pro	2	\$10.13	\$10.13
<b>Total Mensual</b>			<b>\$20.26</b>

Tabla 31: Presupuesto de implementación-licencia Power BI Pro

## Servicios de AWS

### Amazon Simple Storage Service (S3)

Para el cálculo del presupuesto de implementación para el servicio de S3 se hace uso de la calculadora de Amazon desde este sitio <https://calculator.aws/#/>. Este cálculo incluye costos de almacenamiento y costo de solicitudes. A continuación, se detalla el cálculo.



Tiered price for: 10 GB  
10 GB x 0,023 USD = 0,23 USD  
Costo total de la capa = 0,23 USD (coste de almacenamiento en S3 Estándar)  
4000 Solicitudes PUT para almacenamiento de S3 Standard x 0,000005 USD por solicitud = 0,02 USD (coste de solicitudes PUT en S3 Estándar)  
4000 Solicitudes GET en un mes x 0,000004 USD por solicitud = 0,0016 USD (coste de solicitudes GET en S3 Estándar)  
2 GB x 0,0007 USD = 0,0014 USD (coste por las devoluciones de S3 Select)  
5 GB x 0,002 USD = 0,01 USD (coste de escaneados en S3 Select)  
0,23 USD + 0,0016 USD + 0,02 USD + 0,0014 USD + 0,01 USD = 0,26 USD (Total de almacenamiento de S3 Standard, solicitudes de datos, coste de S3 Select)  
**Coste de S3 Estándar (mensual): 0.26 USD**

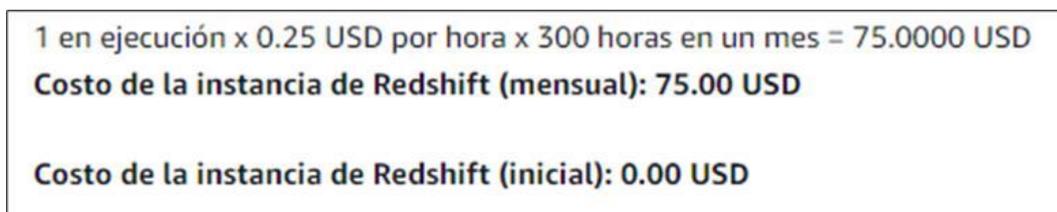
Redondear al valor unitario superior (2.000000) = 2 partes  
**Coste de S3 Estándar (inicial): 0.00 USD**

Costo inicial total: 0.00 USD | Mostrar detalles ▼ | Cancelar  
Costo total mensual: 0,26 USD

Imagen 50: Presupuesto de costo Amazon S3

### Amazon Redshift

Para implementar la solución en Amazon Redshift se estima conveniente utilizar una instancia tipo **dc2.large**, con 2CPU virtuales, 1 nodo y una memoria de 15GiB, además se estima que el uso de la instancia será de 300 horas mensuales. El cálculo se detalla en la siguiente imagen.



1 en ejecución x 0.25 USD por hora x 300 horas en un mes = 75.0000 USD  
**Costo de la instancia de Redshift (mensual): 75.00 USD**

**Costo de la instancia de Redshift (inicial): 0.00 USD**

Imagen 51: Presupuesto de costo Amazon Redshift

Por tanto, considerando los recursos anteriores, el presupuesto mensual de implementación se resume de la siguiente manera:

Recurso	Total
Licencia Power BI Pro	\$20.26
Amazon S3	\$0.26
Amazon Redshift	\$75.00
<b>Total Mensual</b>	<b>\$95.52</b>

Tabla 32: Presupuesto mensual de implementación

### c) Análisis de Resultados

A continuación, se muestra los reportes que dan solución a las necesidades que se buscaron resolver de acuerdo a los requerimientos de la empresa.

#### Reporte de Monto de Compras

Permite visualizar los montos totales de las compras por año, semestre, trimestre y mes o una vista en general, como el usuario desee implementar los filtros de fecha.

#### Monto de Compras

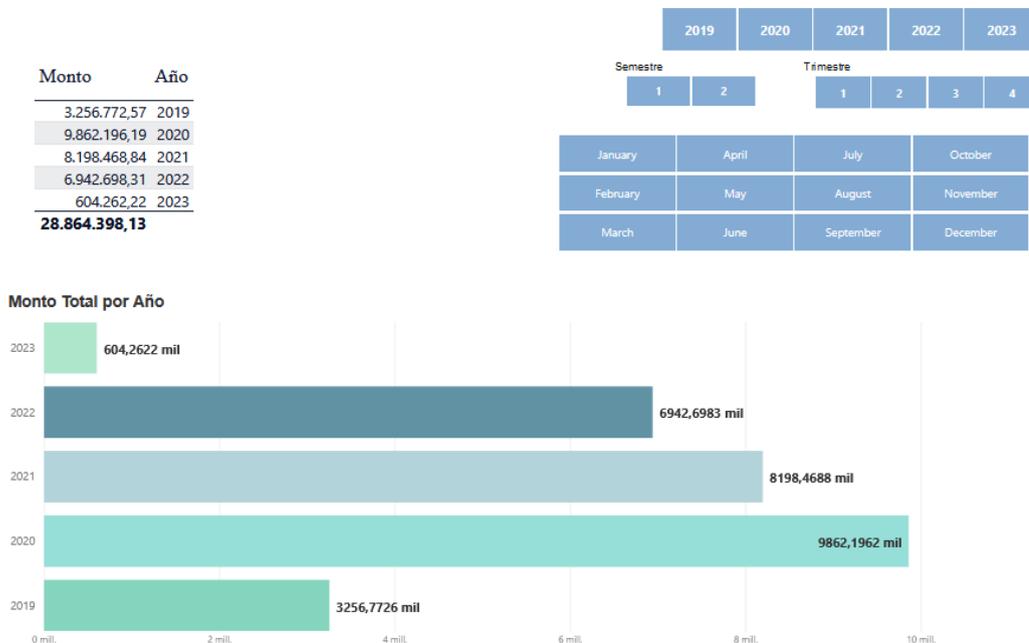


Imagen 52: Informe de Monto de Compras

## Reporte Cantidad de Productos por Proveedor

Presenta la cantidad de productos comprados por proveedor, asimismo permite filtrar por medio de año/semestre/trimestre/mes. Brindando la facilidad de filtrar por producto, obteniendo la cantidad adquirida de ese producto por proveedor.

### Cantidad de Productos Comprados por Proveedor

Proveedor	Cantidad
A. Datum Corporation	7602
International Equipment	5979
Illinois Tech & Solutions	5439
Proseware, Inc.	5080
Electronic Mike Co.	4402
Integrated PC Products	4350
Business Equipment Center	4263
Anderson's Electronics	4202
Electronic Supplies	4003
Litware, Inc.	3771
Trey Research	779
SUPERSALES INC.	500
<b>Total</b>	<b>50370</b>

Producto

2019 2020 2021 2022 2023

Semestre:   Trimestre:

January	April	July	October
February	May	August	November
March	June	September	December

#### Cantidad de Productos por Proveedor

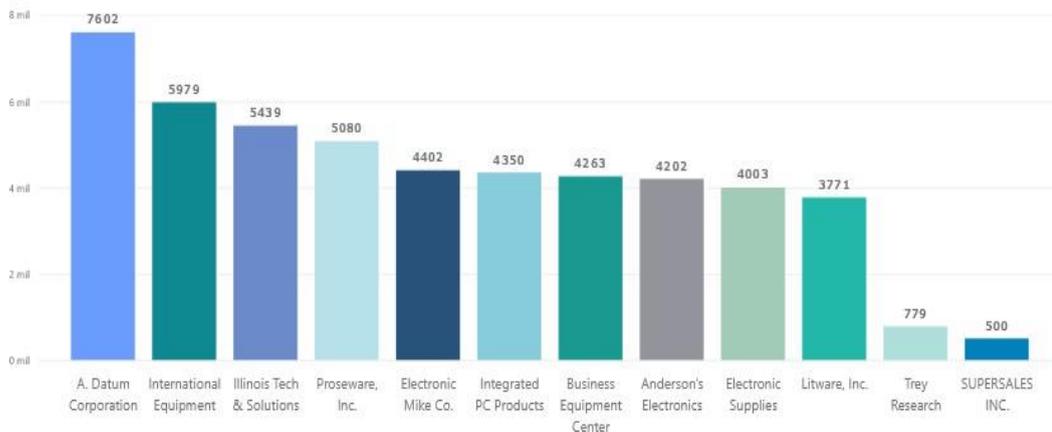


Imagen 53: Informe de Cantidad de Productos por Proveedor

## Reporte Cantidad de Productos Comprados

Genera el producto más comprado, asimismo el que menos se solicitó comprar, esto se puede observar por año, semestre, trimestre, y mes. Asimismo, se puede filtrar por proveedor cual de estos fue más y menos comprado a cada uno de ellos. Además, brinda la tabla con todas las cantidades compradas por producto.

←

Producto	Cantidad
Seagate ST2000DM006	1284
Seagate ST3000DM008	977
Samsung MZ-75E250B/AM	950
Seagate ST1000DM010	863
Supermicro MBD-X10DAX	821
Western Digital WD2500AAJS	785
Western Digital WDS250G1B0A	737
Western Digital WD2500AVVS	669
Western Digital WD2003FZEX	609
SanDisk SDSSDA-240G-G26	601
AMD 100-5056062	548
SanDisk SDSSDA-120G-G26	542
Intel DP35DPM	518
AMD 100-505989	516
MSI X299 GAMING PRO CARBON AC	512
Gigabyte GV-N108TAORUSX W-11GD	502
AData ASU800SS-128GT-C	499
Western Digital WDS256G1W0C	498
Total	50370

### Cantidad de Productos Comprados

Proveedor

Todas

2019	2020	2021	2022	2023
Semestre		Trimestre		
1	2	1	2	3
4				
January	April	July	October	
February	May	August	November	
March	June	September	December	

Producto Mas Comprado

Producto ● Seagate ST2000DM006

Seagate ST2000DM006

1284

Producto Con Menos Compras

Producto ● Corsair Vengeance LPX T8

Corsair Vengeance LPX T8

22

Imagen 54: Informe de Cantidad de Productos Comprados

## Reporte Monto por Producto y Proveedor

Permite visualizar el monto total por producto y proveedor, filtrando por medio de ellos y la fecha que el usuario elija (Año/Semestre/Trimestre/Mes).



Imagen 55: Informe de Monto por Producto y Proveedor

## Reporte Cantidad de Compras

Genera la cantidad de compras que se realizaron por fecha (Año/Semestre/Trimestre/Mes). Además permite filtrar por proveedor mostrando la cantidad de compras por Proveedor.

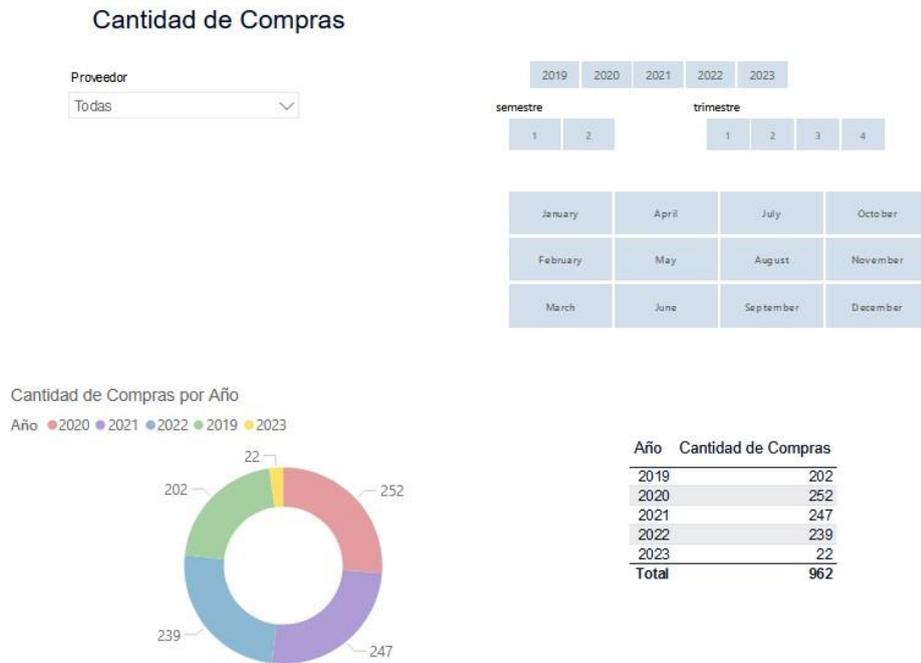


Imagen 56: Informe de Cantidad de Compras

## Conclusiones

Para la realización de este proyecto fue necesario conocer el proceso de compras que la empresa Capital General utiliza, y cuáles son los puntos clave de este, para entender y hacer el respectivo planteamiento de las necesidades para la mejora de esta área, fue preciso hacer el diseño dimensional que daría respuesta a los requerimientos que facilitarían el análisis de la información gestionada a través de ERP NEXT.

Fue una parte fundamental el conocer y definir las métricas los cuales son factores cuantificables que nos permiten analizar de mejor manera los procesos operativos de la empresa, dándonos información específica sobre los distintos datos obtenidos.

Las tareas ETL son parte vital de una solución de un Data Warehouse, ya que nos permite obtener los datos desde nuestro origen de datos, realiza la limpieza de datos correspondiente y carga en las distintas herramientas de almacenamiento y visualización.

Es indispensable asegurar la integridad de los datos y el acceso a ellos en cualquier solución, en soluciones en la nube como lo es con los servicios de Amazon se puede coordinar fácilmente por medio de usuarios, roles y políticas.

Herramientas de análisis empresariales tales como Power BI son cruciales para la presentación de la información transformada, agregando que permite la conexión de forma rápida a recursos tales como Amazon Redshift, y permite la creación de dashboard e informes para una vista de los datos que sean fácil para el usuario el análisis y el entendimiento de los datos.

A través de la solución la empresa Capital General obtuvo de manera grafica información ya transformada que le permitirá analizarla y tomar decisiones que mejoren el proceso de compras en la compañía.

## Recomendaciones

Desde la perspectiva de desarrollo de software, se identifica una oportunidad clave para Capital General al integrar un Data Warehouse en su infraestructura tecnológica. Aunque ERP NEXT es eficaz para la gestión operativa, hay una brecha en la capacidad analítica que un Data Warehouse puede llenar adecuadamente.

Al centralizar los datos de compras y proveedores, un Data Warehouse ofrece una única fuente de verdad, permitiendo la ejecución de consultas complejas y la generación de informes detallados. Este enfoque mejoraría significativamente la visibilidad de las operaciones de compra y proporcionaría a la gerencia información crucial para el análisis estratégico.

Los informes generados podrían incluir estadísticas importantes como volumen de compras, distribución de gastos por proveedor, totales acumulados y tendencias de compra, facilitando una visión global del rendimiento del departamento de compras. Al final, la implementación de esta solución conduciría a una toma de decisiones más informada, impulsando así la eficiencia y competitividad de la empresa en su sector.

## Bibliografía

Amazon Elastic Compute. (s.f.). *What is EC2*. Obtenido de AWS Documentation: <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/concepts.html>

Amazon Redshift. (s.f.). *Amazon Redshift*. Obtenido de AWS Documentation: <https://docs.aws.amazon.com/redshift/latest/gsg/new-user-serverless.html>

Amazon Simple Storage Service. (s.f.). *What is Amazon S3?* Obtenido de AWS Documentation: <https://docs.aws.amazon.com/AmazonS3/latest/userguide/Welcome.html>

AWS Pricing Calculator. (s.f.). *AWS Pricing Calculator user guide*. Obtenido de AWS Documentation: [https://docs.aws.amazon.com/es\\_es/pricing-calculator/latest/userguide/getting-started.html](https://docs.aws.amazon.com/es_es/pricing-calculator/latest/userguide/getting-started.html)

ERPNext. (s.f.). *ERPNext User Guide*. Obtenido de ERPNext Documentation: <https://docs.erpnext.com/docs/user/manual/en/introduction>

Forum. (s.f.). *ERPNext Production\_v14.8.0 ova download*. Obtenido de Frappe Forum: <https://discuss.frappe.io/t/erpnext-production-v14-8-0-ova-download-link/96926>

Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The definitive guide to Dimensional Modeling, Third Edition*.

Power BI. (s.f.). *Power BI Desktop*. Obtenido de Power BI Documentation: <https://learn.microsoft.com/en-us/power-bi/>

Talend. (s.f.). *Talend Open Studio* . Obtenido de Talend Open Studio Documentation: <https://help.talend.com/r/en-US/8.0/open-studio-user-guide>

## Anexos

### Anexo 1: Diagrama de proceso de Orden de Compras.

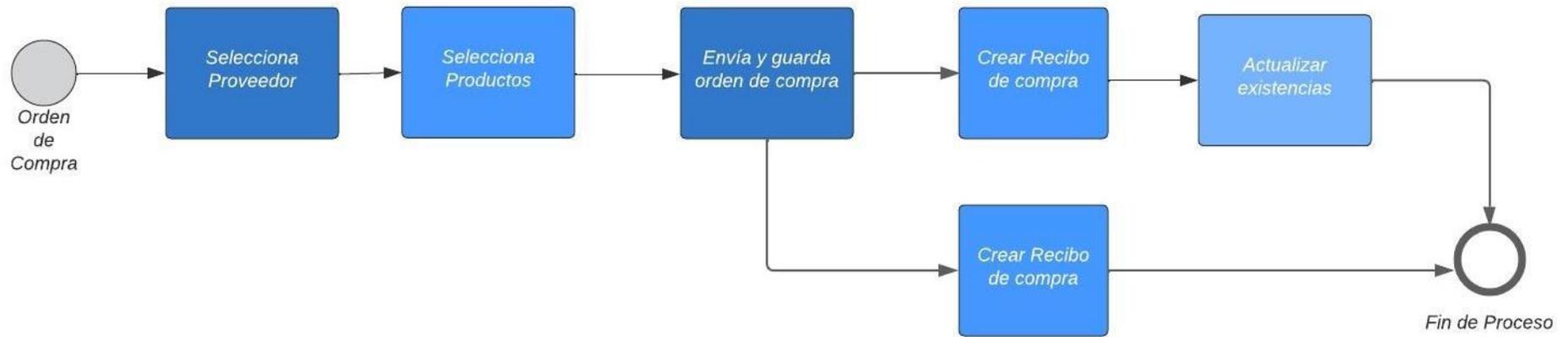
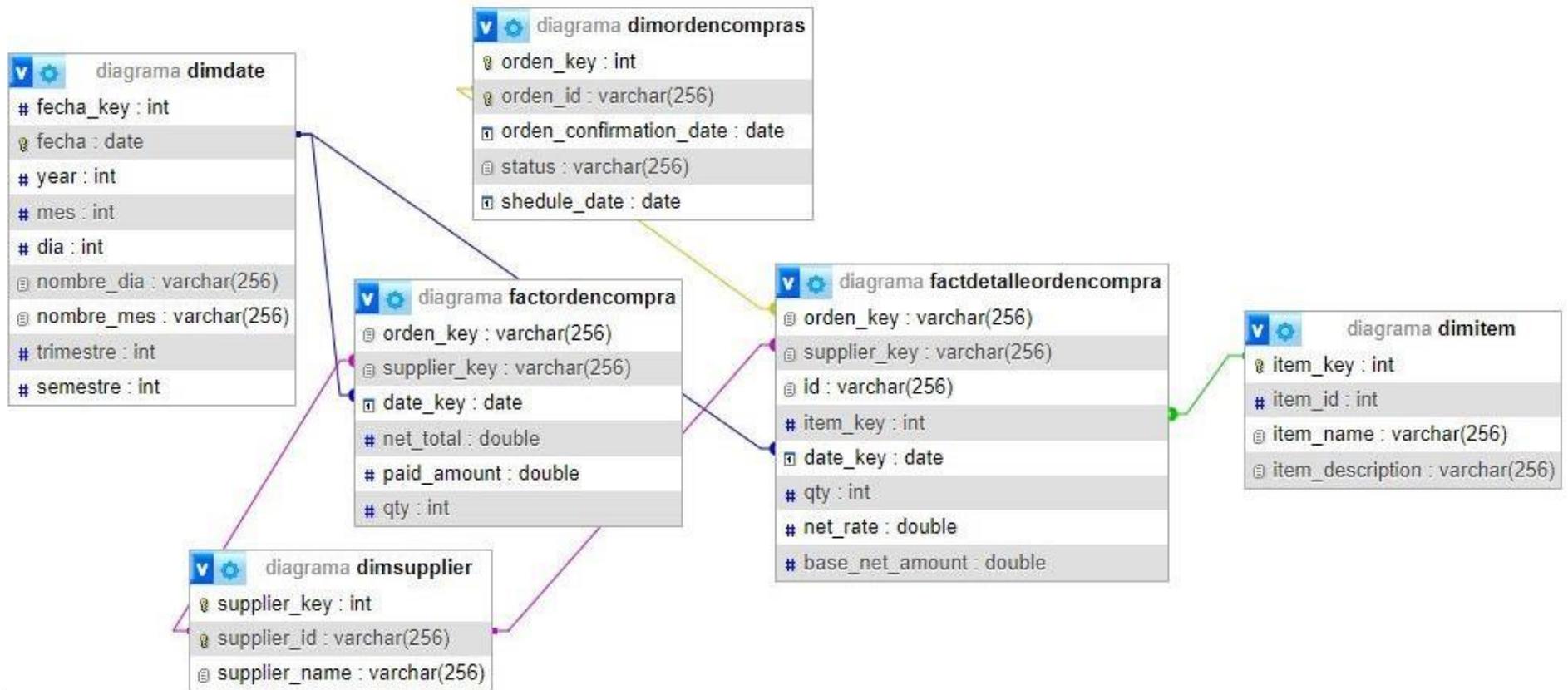


Imagen 57: Proceso de Orden de Compra

## Anexo 2: Diagrama del Modelo Dimensional Data Warehouse



sola

Imagen 58: Modelo Dimensional de Data Warehouse

### Anexo 3: Diagrama Gantt De Actividades

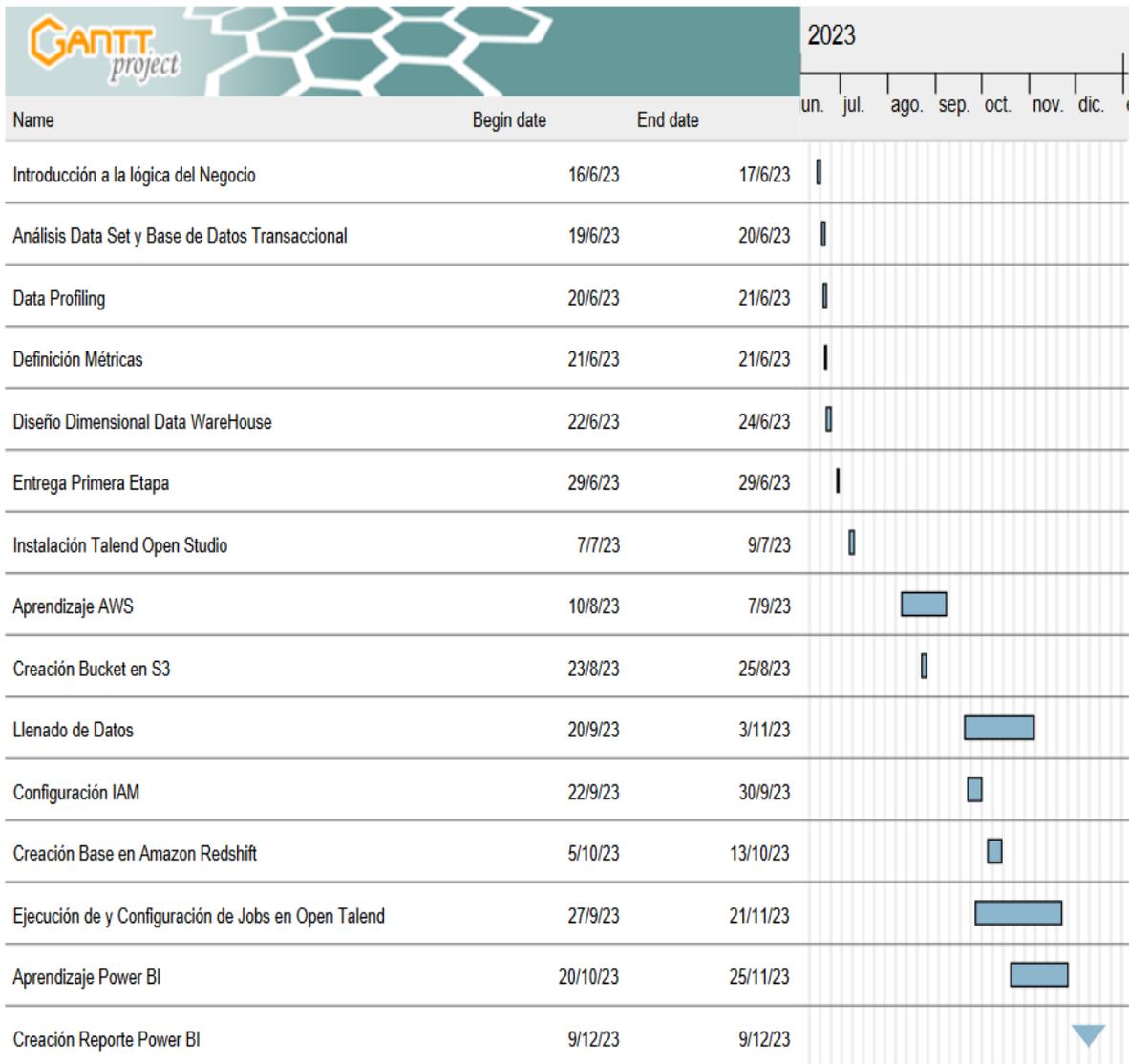


Imagen 59: Diagrama de Gantt - Cronograma de actividades