

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



CURSO DE ESPECIALIZACIÓN EN INGENIERÍA DE DATOS
**ANÁLISIS DE INVENTARIO PARA LA EMPRESA DE
LOGÍSTICA EL SALVADOR COURIER QUE POSEE 3
BODEGAS FÍSICAS ADMINISTRADAS A TRAVÉS DEL
SOFTWARE PARTKEEPR**

PRESENTADO POR:

JORGE GEOVANI HERNÁNDEZ AMAYA

AXEL OMAR HERNÁNDEZ QUINTANILLA

PARA OPTAR AL TÍTULO DE:

INGENIERO DE SISTEMAS INFORMÁTICOS

CIUDAD UNIVERSITARIA, MARZO DE 2023

UNIVERSIDAD DE EL SALVADOR

RECTOR:

MSC. ROGER ARMANDO ARIAS ALVARADO

SECRETARIO GENERAL:

ING. FRANCISCO ANTONIO ALARCÓN SANDOVAL

FACULTAD DE INGENIERÍA Y ARQUITECTURA

DECANO:

PHD. EDGAR ARMANDO PEÑA FIGUEROA

SECRETARIO:

ING. JULIO ALBERTO PORTILLO

ESCUELA DE INGENIERÍA DE SISTEMAS INFORMÁTICOS

DIRECTOR:

ING. RUDY WILFREDO CHICAS VILLEGAS

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:

**ANÁLISIS DE INVENTARIO PARA LA EMPRESA DE
LOGÍSTICA EL SALVADOR COURIER QUE POSEE 3
BODEGAS FÍSICAS ADMINISTRADAS A TRAVÉS DEL
SOFTWARE PARTKEEPR**

Presentado por:

JORGE GEOVANI HERNÁNDEZ AMAYA

AXEL OMAR HERNÁNDEZ QUINTANILLA

Curso de Especialización Aprobado por:

Docente Asesor:

ING. RENÉ FABRICIO QUINTANILLA GÓMEZ

SAN SALVADOR, MARZO DE 2023

Trabajo de Graduación Aprobado por:

Docente Asesor:

Ing. René Fabricio Quintanilla Gómez

ÍNDICE

Introducción	i
Capítulo I: Especificación de proyecto	1
a. Objetivos	1
b. Situación actual.....	2
i. Antecedentes.....	2
ii. Descripción del problema	3
iii. Planteamiento del problema	3
c. Alcances	4
d. Justificación	5
e. Cronograma de actividades.....	6
f. Presupuesto.....	7
Capitulo II: Análisis y diseño de la propuesta de solución	8
a. Metodología de trabajo.....	8
b. Descripción de la propuesta de solución	9
Modelo de snapshot diario de inventario.....	9
Modelo de transacciones de inventario.....	11
c. Descripción de la tecnología a utilizar	17
d. Diagrama arquitectónico de la solución.....	19
e. Descripción de cada componente de la solución.....	20
Capitulo III: Estrategia de implementación de propuesta de solución.....	22
a. Estrategia de implementación	22
b. Presupuesto de implementación	22
c. Análisis de resultados	23
Conclusiones y recomendaciones.....	25
Bibliografía.....	26
Anexos	27

INTRODUCCIÓN

En el presente trabajo se pretende mostrar la aplicación de la ingeniería de datos a un caso práctico, específicamente al proceso de negocio relacionado con el análisis de inventario de productos. Para este efecto, se ha tomado como referencia la empresa El Salvador Courier la cual es una empresa de logística que administra tres bodegas de productos siendo estos principalmente piezas electrónicas. El Salvador Courier cuenta con un software denominado PartKeepr el cual le permite almacenar los datos relacionados a las transacciones de inventario que se realizan a diario. Este software PartKeepr y principalmente los datos contenidos en su base de datos se convertirán en la fuente principal que alimentará a la solución que en este documento se describe.

Antes de entrar en detalle a la solución propuesta, se hace necesario describir la situación actual que enfrenta la empresa, es por eso que se incluye en el primer capítulo *Especificación del proyecto* los correspondientes antecedentes de la empresa para entrar en materia de las funciones principales que la empresa lleva a cabo. Asimismo, luego de tener una vista general de la empresa se empieza a identificar cuál es la problemática a resolver y cómo esto ayudaría a crear valor en la empresa.

La solución propuesta en este documento está basada tomando en cuenta la literatura presentada en el libro denominado *The Data Warehouse Toolkit. The Definitive Guide to Dimensional Modeling* de los autores Ralph Kimball y Margy Ross. Este libro presenta las pautas para un buen diseño de datawarehouse para el proceso de negocio *Inventario*. En el capítulo II de este documento se describe cuáles son los modelos de análisis de inventario que se han tomado en cuenta para el diseño de la solución aquí presentada. Inmediatamente después de describir la metodología de trabajo utilizada, se presenta el diseño de datawarehouse propuesto..

Es muy importante destacar también el papel que juega el uso de las herramientas tecnológicas a disposición para la creación de la solución por lo cual en este documento se describe cómo Amazon Web Services es utilizado como el principal lugar de almacenamiento de los datos a través del servicio Amazon S3 y la posterior carga de los datos mediante herramientas ETL a una estructura de datawarehouse en Amazon Redshift. Los datos aquí almacenados son consumidos finalmente por la aplicación de visualización Power BI, la cual se convierte en la principal fuente de información para usuarios finales.

CAPÍTULO I: ESPECIFICACIÓN DE PROYECTO

A. OBJETIVOS

Objetivo general

Crear un Datawarehouse a partir de la base de datos transaccional del software PartKeepr que facilite a la empresa “El Salvador Courier” la creación de los reportes que necesita.

Objetivos específicos

- Seleccionar las tablas relevantes de la base de datos de PartKeepr que serán utilizadas para construir el datawarehouse.
- Crear el modelado dimensional del datawarehouse con las tablas seleccionadas.
- Crear un volcado de todas las tablas seleccionadas en archivos .csv.
- Crear un “bucket” dentro de Amazon S3.
- Crear una estructura dentro del “bucket” de Amazon S3 donde se almacenen los archivos extraídos de la base de datos transaccional y los archivos procesados por el datawarehouse.
- Implementar en Amazon Redshift el modelo dimensional del datawarehouse.

B. SITUACIÓN ACTUAL

i. Antecedentes

La empresa “El Salvador Courier” es una empresa de logística creada en 2009, con sede principal en San Salvador, donde está ubicada su bodega principal, así como también 2 bodegas más con ubicación en La Libertad y Santa Ana, dando un total de 3 bodegas físicas para administrar el inventario de sus clientes.

Esta empresa maneja inventario específicamente en el rubro de electrónica, se administran piezas como semiconductores, circuitos integrados, capacitores, conectores, etc. Para llevar la administración de este inventario la empresa El Salvador Courier hace uso del software PartKeepr, el cual cumple con las necesidades de la empresa.

La empresa utiliza el software PartKeepr para poder realizar los siguientes movimientos de inventario:

Cada vez que una cantidad de cierto producto ingresa a la bodega, este movimiento es registrado como “entrada” aumentando a la vez la cantidad disponible de dicho producto en ese momento.

Cada vez que una cantidad de cierto producto sale de bodega, este movimiento es registrado como “salida” disminuyendo a la vez la cantidad disponible de dicho producto en ese momento.

Adicionalmente, la empresa administra proyectos. Estos proyectos están conformados por componentes que ya están almacenados en bodega. Para poder retirar los componentes asignados a un proyecto se debe crear un reporte de proyecto. Al realizar esta acción el nivel de inventario también es actualizado.

Se realiza inspección de inventario cada 3 meses para corroborar que los datos almacenados en PartKeepr sean correctos, ejecutando las correcciones de ser necesarias.

ii. Descripción del problema

La generación de reportes es esencial en todas las empresas, ya que permiten a los tomadores de decisiones ver el panorama general sobre el desempeño de la empresa en un período de tiempo determinado, o el desempeño que ha tenido alguno de sus productos o servicios, entre otras interrogantes. Algunas de las preguntas que los reportes nos permiten responder son: ¿Cuánto de x producto se ha vendido en un período de tiempo? ¿Cuáles han sido las ganancias totales por día, mes y/o año? ¿Qué porcentaje de producto resultó defectuoso?, etc.

Además de responder estas preguntas, los reportes presentan la información de manera compacta, con elementos como gráficos de pastel o de barras, líneas de tiempo, tablas resumen y entre otros, que permiten que la información sea entendida fácilmente.

Actualmente, la empresa "El Salvador Courier" no posee los medios para generar reportes de una manera práctica, ya que el software PartKeepr no provee de funciones que permitan generarlos. La solución actual a esta problemática es crear volcados de la base de datos a un archivo .csv para poder ser manipulado en Microsoft Excel e ir creando los reportes de manera manual.

Es por esto que la empresa ha tomado la decisión de crear un DataWarehouse que facilite la creación de los reportes que necesita, esto apoyado del software Microsoft PowerBI y los servicios en la nube de Amazon Web Services, concretamente Amazon S3 y Amazon Redshift. Este DataWarehouse será creado a partir de la base de datos transaccional de PartKeepr, pero optimizado ocupando solo las tablas relevantes para la creación de los reportes.

iii. Planteamiento del problema

La creación de un DataWarehouse hecho a partir de la base de datos transaccional del software PartKeepr utilizando los servicios en la nube de Amazon Web Services, que facilite a la empresa "El Salvador Courier" la generación de reportes que responda a sus necesidades.

C. ALCANCES

- Creación de un datawarehouse a partir de la base de datos transaccional del software PartKeepr que se ajuste a las necesidades de la empresa y permita la generación de los reportes que necesita.
- Medición del porcentaje de rotación de inventario diaria para un producto y bodega en específico.
- Medición del porcentaje de rotación de inventario durante un período de tiempo para un producto y bodega en específico.
- Cantidad de días de suministro disponibles para un producto y bodega en un momento determinado.
- Costo total de los productos despachados diariamente en una bodega en específico.
- Conteo de la entrada de unidades de un producto en una bodega específica durante un período de tiempo.
- Conteo de la salida de unidades de un producto en una bodega específica durante un período de tiempo.
- Conteo de la cantidad de unidades de un determinado producto que estaba almacenado en una bodega específica y que haya sido utilizado como parte de un proyecto durante un período de tiempo.
- Conteo de la cantidad de unidades de un determinado producto que se encontraron dañados o defectuosos en el desarrollo de inspecciones de inventario durante un periodo de tiempo.

D. JUSTIFICACIÓN

El presente trabajo tiene como objetivo facilitar la creación de reportes mediante la implementación de un datawarehouse que se ajuste a las necesidades de la empresa. El motivo de su realización surge de la carencia de empresas a nivel nacional que emplean este tipo de soluciones. Es un campo que aún no ha sido muy explorado en el país, por lo que es una buena oportunidad para mostrar los beneficios que estas soluciones traen a las empresas, impulsando su uso.

Adicional, la implementación de este tipo de soluciones en la nube trae como beneficio la reducción de costos en equipo de infraestructura, ya que la solución se despliega sobre la infraestructura en la nube, solo se incurren en costos de uso de estos servidores, sea costo por hora, por día o por mes. Además, podemos aumentar o reducir la capacidad de procesamiento según sea requerido y se logra una mayor disponibilidad de la solución.

E. CRONOGRAMA DE ACTIVIDADES

Actividad	Duración	Fecha inicio	Fecha fin
Levantamiento de requerimientos	5 días	01/09/2022	07/09/2022
Análisis de base de datos transaccional de PartKeepr	3 días	08/09/2022	12/09/2022
Creación del modelo dimensional del datawarehouse	7 días	13/09/2022	22/09/2022
Volcado de base de datos transaccional de PartKeepr	1 día	23/09/2022	23/09/2022
Creación de bucket y estructura de carpetas en Amazon S3	1 día	26/09/2022	26/09/2022
Creación de procesos ETL para subida de archivos a capa "raw" en bucket de Amazon S3	1 día	26/09/2022	26/09/2022
Implementación del modelo dimensional en Amazon Redshift	5 días	27/09/2022	03/10/2022
Creación de procesos ETL para transformación de datos y subida de archivos a capa "staging" en bucket de Amazon S3	20 días	04/10/2022	01/11/2022
Creación de procesos ETL para transformación de datos y subida de archivos a capa "presentation" en bucket de Amazon S3	2 días	03/11/2022	04/11/2022
Creación de procesos ETL para llenar el datawarehouse en Amazon Redshift a partir de la capa de presentación del bucket en Amazon S3	7 días	07/11/2022	15/11/2022
Elaboración de reportes en Microsoft PowerBI	7 días	16/11/2022	24/11/2022

F. PRESUPUESTO

A continuación, se detalla el presupuesto para el desarrollo del proyecto. Al ser una implementación en la nube, y al ya contar con el equipo necesario para su desarrollo, los únicos costos incurridos son por el uso de los servicios en la nube y pago de licencias. El costo de los servicios en la nube varía con el tiempo. Además, el desarrollo del proyecto se realizó de manera remota, por lo que no se incurrió en gastos de transporte y alimentación.

Nombre	Monto
Amazon Redshift, nodo dc2.large	\$0.25 nodo/hora, \$109.50 nodo/mes, \$1,314 nodo/año
Amazon S3	\$12.10 al mes, \$145.20 al año
Suscripción Microsoft PowerBI	\$30 al mes, \$360 al año
2 desarrolladores a tiempo completo	\$900 al mes c/u

CAPITULO II: ANÁLISIS Y DISEÑO DE LA PROPUESTA DE SOLUCIÓN

A. METODOLOGÍA DE TRABAJO

Para la construcción de la solución se hizo uso del contenido expuesto en el libro *The Data Warehouse Toolkit. The Definitive Guide to Dimensional Modeling* el cual es presentado por los autores Ralph Kimball y Margy Ross.

Este libro presenta los principios básicos para el buen diseño de un datawarehouse, explicando detalladamente los conceptos fundamentales que entran en juego en este diseño como lo son los conceptos de tablas de hechos y tablas de dimensión principalmente. Antes de entrar en detalle de la solución es necesario recordar que las tablas de dimensión son las que darán contexto a la información almacenada mientras que las tablas de hecho son las que guardarán los registros de los datos que pueden ser medidos y con los cuales se pueden tomar decisiones,

En este trabajo, se hace uso principalmente del capítulo 4 del libro mencionado anteriormente. El capítulo 4 es titulado *Inventory* y presenta las pautas a considerar para un buen diseño de datawarehouse cuando se trata de analizar los datos almacenados en un sistema de inventario.

Se proponen en el libro de Kimball tres modelos de datawarehouse de inventario: el modelo de snapshot periódico, el modelo de transacciones y el modelo de snapshot acumulativo. El modelo de snapshot periódico está diseñado para almacenar periódicamente un registro por cada producto en el inventario por cada tienda en que se tiene el producto. Cada registro debe guardar como elemento fundamental la cantidad de producto existente al final del día o al momento en que se decida guardar dicha información. Adicionalmente se puede guardar el costo unitario de dicho producto a la fecha en que se guarda la información, así como la cantidad despachada en ese día. Con esta información en el datawarehouse puede calcularse la rotación de inventario y los días de suministro de inventario tal y como lo menciona el libro de Kimball en la página # 115 de la tercera edición.

El segundo modelo propuesto es el modelo de transacciones de inventario el cual tiene por objetivo principalmente guardar los datos pertenecientes a los movimientos de inventario que se llevan a cabo reflejando así tanto los registros de entradas como salidas de productos que

son realizadas y de esta manera poder analizar cuál es el flujo de un cierto producto en inventario y tomar las decisiones correspondientes.

El tercer modelo que se propone en el libro de Kimball es el modelo de snapshot acumulativo, el cual tiene como propósito guardar un registro por cada lote de producto y sobre el mismo registro llevar un seguimiento de las diferentes transacciones que se van realizando de ese lote de producto. No se entrará en detalle sobre este modelo ya que para la solución propuesta en este trabajo solamente se contemplarán el modelo de snapshot periódico y el modelo de transacciones de inventario.

B. DESCRIPCIÓN DE LA PROPUESTA DE SOLUCIÓN

A continuación se muestra detalladamente la propuesta de datawarehouse diseñada que se utilizará para realizar el análisis de inventario de la empresa Courier. Como se mencionó en el apartado anterior, la siguiente propuesta de solución incluye dos modelos de análisis de inventario: el modelo de snapshot periódico o snapshot diario de inventario y el modelo de transacciones de inventario. Para ambos modelos se describen los pasos de diseño realizados como lo son la selección del proceso de negocio, definición de granularidad, identificación de dimensiones, identificación de métricas, así como el esquema resultante el cual muestra las relaciones entre dimensiones y la respectiva tabla de hechos para cada modelo presentado.

Modelo de snapshot diario de inventario

Selección del proceso de negocio

Snapshot diario del inventario en bodegas

Definición de granularidad

Una fila por snapshot diario del nivel de inventario para cada producto (componente) en cada bodega

Identificación de dimensiones

Producto: productoKey, idProducto, nombreProducto, descripcion, comentarios, unidadMedida, abreviaturaUnidadMedida, categoria, subcategoria, nombreHuella, categoriaHuella, cantidadMinimaPermitida, estado, fechaCreacion, codigoInterno.

Fecha: fechaKey, fechaCompleta, nombreDia, abreviaturaDia, numeroDiaEnMes, nombreMes, abreviaturaMes, numeroMesEnAnio, anio, anioMes, numeroSemanaEnAnio.

Bodega: bodegaKey, idBodega, nombreBodega.

Identificación de métricas:

- cantidadDisponible
- cantidadDespachada

Para obtener la rotación de inventario diaria de un producto se dividirá la cantidad despachada entre la cantidad disponible de ese día.

Para obtener la rotación de inventario de un producto durante un periodo se debe dividir la **suma de cantidades despachadas** del periodo entre el **promedio de cantidades disponibles diariamente** durante ese periodo.

Para obtener la cantidad de días de suministro de un producto se debe dividir la **cantidad disponible que se tiene al final** de un periodo entre el **promedio de cantidades despachadas diariamente** durante ese periodo.

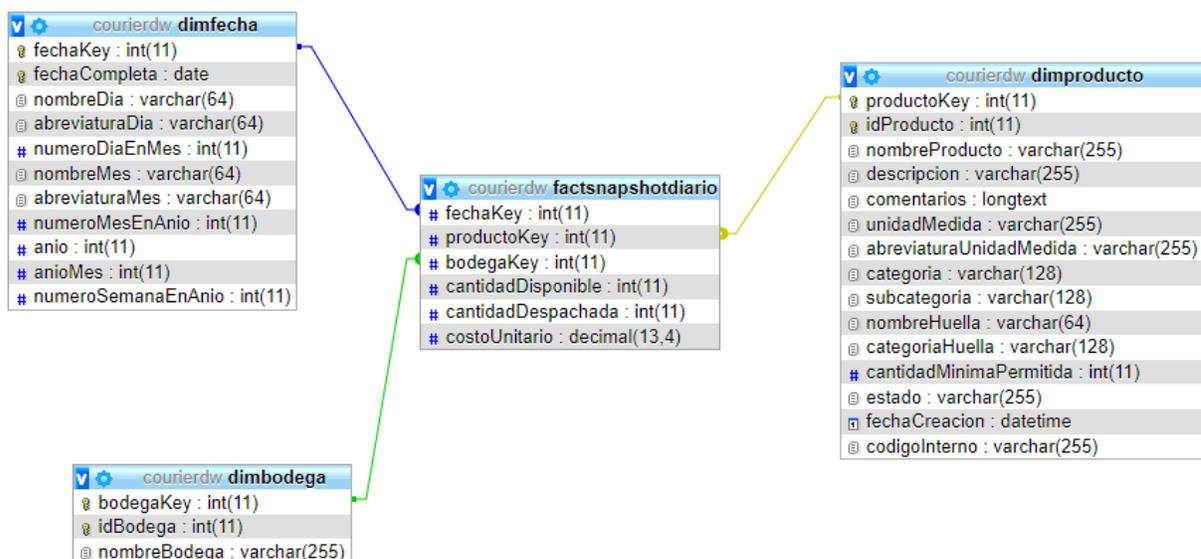


Ilustración 1. Diagrama de datawarehouse del modelo de snapshot diario de inventario.

Modelo de transacciones de inventario

Selección del proceso de negocio

Transacciones de inventario

Definición de granularidad

Una fila por cada transacción de producto en inventario (entrada a bodega, salida de bodega, asignación de productos a proyectos, inspección).

Identificación de dimensiones

Producto: productoKey, idProducto, nombreProducto, descripcion, comentarios, unidadMedida, abreviaturaUnidadMedida, categoria, subcategoria, nombreHuella, categoriaHuella, cantidadMinimaPermitida, estado, fechaCreacion, codigoInterno.

Fecha: fechaKey, fechaCompleta, nombreDia, abreviaturaDia, numeroDiaEnMes, nombreMes, abreviaturaMes, numeroMesEnAnio, anio, anioMes, numeroSemanaEnAnio.

Bodega: bodegaKey, idBodega, nombreBodega.

TipoTransaccion: tipoTransaccionKey, nombreTipoTransaccion.

Identificación de métricas

- cantidadEnUnidades
- costoUnitario
- cantidadEnDinero
- conteo

Para obtener el costo total de los productos despachados diariamente en una bodega en específico se multiplicará la **cantidad en unidades** por el **costo unitario** de las transacciones identificadas como “salida” de ese día.

Para obtener el costo total de los productos despachados durante un periodo de tiempo en una bodega en específico se multiplicará la **cantidad en unidades** por el **costo unitario** de las transacciones identificadas como “salida” de ese periodo y luego se obtendrá la suma de dichas multiplicaciones.

Para obtener la cantidad de unidades de un determinado producto que entraron a una bodega durante un periodo de tiempo en específico se sumará las **cantidades en unidades** de las transacciones identificadas como “entrada” que ocurrieron en el periodo.

Para obtener la cantidad de unidades de un determinado producto que salieron de una bodega durante un periodo de tiempo en específico se sumará las **cantidades en unidades** de las transacciones identificadas como “salida” que ocurrieron en el periodo.

Para obtener la cantidad de unidades de un determinado producto que ha sido utilizada en proyectos se sumará **las cantidades en unidades** de las transacciones identificadas como “proyecto” que ocurrieron en el periodo.

Para obtener la cantidad de unidades de un determinado producto que se encontraron dañados o defectuosos en el desarrollo de inspecciones de inventario durante un periodo de tiempo en específico se sumará las **cantidades en unidades** de las transacciones identificadas como “inspección” que ocurrieron en el periodo.

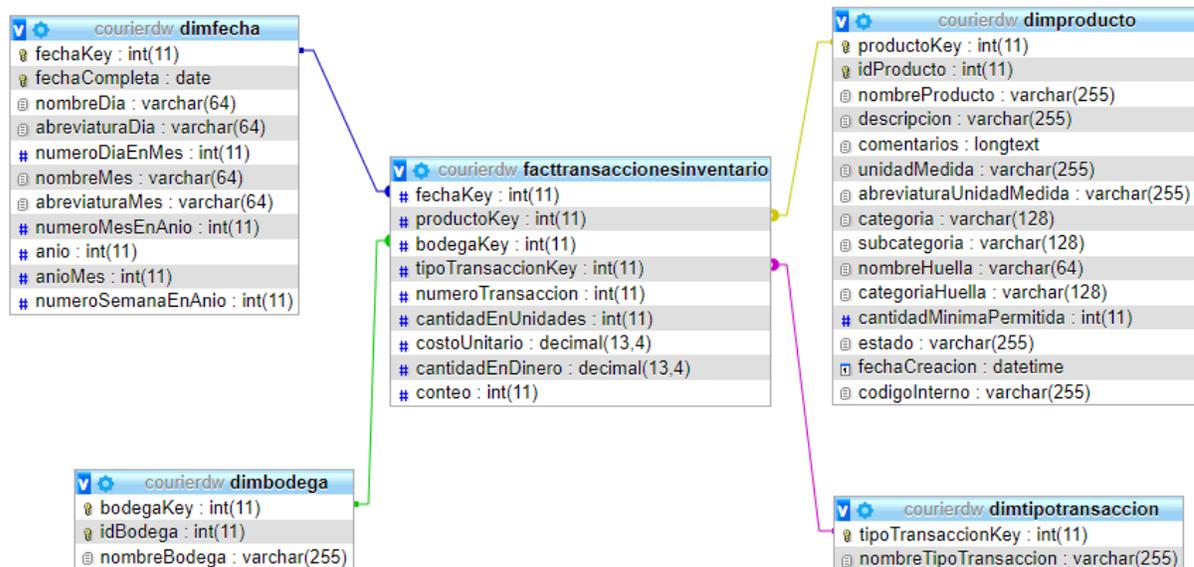


Ilustración 2. Diagrama de datawarehouse del modelo de transacciones de inventario.

Diccionario de datos

Dimensión DimProducto

Almacena los datos de los productos que se tienen en el inventario.

Columna	Tipo	Nulo	PK	Enlaces a
productoKey	int(11)	No	PK	
Llave subrogada				
idProducto	int(11)	No		
Llave de negocio				
nombreProducto	varchar(255)	No		
Nombre del producto				
descripcion	varchar(255)	No		
Breve descripción del producto				
comentarios	longtext	No		
Comentarios adicionales acerca del producto				
unidadMedida	varchar(255)	No		
Unidad en que se mide el producto (ej: pieza, centímetro, mililitro)				
abreviaturaUnidadMedida	varchar(255)	No		
Abreviatura de la unidad en que se mide el producto (ej: pcs, cm, ml)				
categoria	varchar(128)	No		
Nombre de la categoría a la que pertenece el producto				
subcategoria	varchar(128)	No		
Nombre de la subcategoría a la que pertenece el producto				
nombreHuella	varchar(64)	No		
Nombre del tipo de huella del producto (ej: CBGA-32, PBGA-119)				
categoriaHuella	varchar(4)	No		
Nombre de la categoría de huella del producto (ej: CBGA, PBGA)				
cantidadMinimaPermitida	int(11)	No		
Cantidad mínima de producto que debe haber en el inventario				
estado	varchar(16)	No		

Estado del producto en el inventario (Activo, Inactivo)				
fechaCreacion	datetime	No		
Fecha y hora en que se registró el producto en la base de datos				
codigoInterno	varchar(255)	No		
Código identificador del producto dentro del inventario				

Dimensión DimBodega

Almacena los datos de las bodegas en las que se almacenan productos.

Columna	Tipo	Nulo	PK	Enlaces a
bodegaKey	int(11)	No	PK	
Llave subrogada				
idBodega	int(11)	No		
Llave de negocio				
nombreBodega	varchar(255)	No		
Nombre de la bodega				

Dimensión DimFecha

Almacena un registro de fechas con sus elementos relevantes.

Columna	Tipo	Nulo	PK	Enlaces a
fechaKey	int(11)	No	PK	
Llave subrogada en formato ISO (ej: 20180101 para identificar 01 de enero de 2018)				
fechaCompleta	date	No		
Fecha en el formato dd-mm-YYYY				
nombreDia	varchar(16)	No		
Nombre del día (ej: Lunes)				
abreviaturaDia	varchar(2)	No		
Abreviatura del día (ej: Lu, Ma, Mi)				
numeroDiaEnMes	int(11)	No		

Número del día en el mes (en el rango 1-31)				
nombreMes	varchar(16)	No		
Nombre del mes del año (ej: Enero)				
abreviaturaMes	varchar(4)	No		
Abreviatura del nombre del mes (ej: Ene, Feb)				
numeroMesEnAnio	int(11)	No		
Número del mes en el año (en el rango 1-12)				
anio	int(11)	No		
Año (ej: 2018)				
anioMes	int(11)	No		
Número del año seguido del número del mes (ej: 201801 para identificar enero de 2018)				
numeroSemanaEnAnio	int(11)	No		
Número de la semana en el año a la que pertenece la fecha				

Dimensión DimTipoTransaccion

Almacena los datos de los tipos de transacciones de inventario que se realizan en las bodegas.

Columna	Tipo	Nulo	PK	Enlaces a
tipoTransaccionKey	int(11)	No	PK	
Llave subrogada				
nombreTipoTransaccion	varchar(16)	No		
Nombre del tipo de transacción (Ej: Entrada, Salida, Inspección)				

Tabla de hechos FactSnapshotDiario

Almacena una fila por snapshot diario del nivel de inventario para cada producto en cada bodega.

Columna	Tipo	Nulo	PK	Enlaces a
fechaKey	int(11)	No	PK	DimFecha
Llave foránea a dimensión DimFecha				

productoKey	int(11)	No	PK	DimProducto
Llave foránea a dimensión DimProducto				
bodegaKey	int(11)	No	PK	DimBodega
Llave foránea a dimensión DimBodega				
cantidadDisponible	int(11)	No		
Cantidad disponible del producto en la bodega al momento en que se realiza el snapshot				
cantidadDespachada	int(11)	No		
Cantidad despachada del producto desde la bodega en el día en que se realiza el snapshot				
costoUnitario	decimal(13,4)	No		
Costo unitario en dólares del producto al momento en el que se realiza el snapshot				

Tabla de hechos FactTransaccionesInventario

Almacena una fila por cada transacción de producto en inventario (entrada a bodega, salida de bodega, asignación de productos a proyectos, inspección).

Columna	Tipo	Nulo	PK	Enlaces a
fechaKey	int(11)	No	PK	DimFecha
Llave foránea a dimensión DimFecha				
productoKey	int(11)	No	PK	DimProducto
Llave foránea a dimensión DimProducto				
bodegaKey	int(11)	No	PK	DimBodega
Llave foránea a dimensión DimBodega				
tipoTransaccionKey	int(11)	No	PK	DimTipoTransaccion
Llave foránea a dimensión DimTipoTransaccion				
numeroTransaccion	int(11)	No	PK	
Número identificador de la transacción				
cantidadEnUnidades	int(11)	No		
Cantidad en unidades de productos que entran o salen de bodega				
costoUnitario	decimal(13,4)	No		
Costo unitario del producto en dólares al momento de la transacción				

cantidadEnDinero	decimal(13,4)	No		
Costo total en dólares de los productos que entran o salen de bodega				
conteo	int(11)	No		
Este campo almacena el número 1 por cada transacción y se utiliza para realizar el conteo de las transacciones realizadas durante el día o cualquier periodo en específico.				

C. DESCRIPCIÓN DE LA TECNOLOGÍA A UTILIZAR



Ilustración 3. Logo de Amazon S3.

Amazon S3: Este servicio es conocido también como Amazon Simple Storage Service o Amazon S3. Es un servicio de almacenamiento de objetos en la nube ofrecido en Amazon Web Services. Sus principales características son la escalabilidad, la disponibilidad de datos, la seguridad y el rendimiento. Es muy útil para almacenar lo que es conocido como lago de datos, aplicaciones nativas en la nube e incluso aplicaciones móviles. Además de ser muy rentable es muy fácil de usar y administrar por lo que también se pueden crear diferentes niveles de acceso conocidos como IAM (Identity and Access Management) de acuerdo a los requisitos de la empresa. La capa gratuita de Amazon S3, la cual se ha utilizado para este proyecto, ofrece 5 GB de almacenamiento durante 12 meses. En el apartado *Descripción de cada componente de la solución* de este capítulo se explicará detalladamente cómo S3 ha sido de utilidad en este proyecto.

Redshift: Es un servicio en la nube de Amazon Web Services y su principal utilidad es el análisis de datos estructurados y semiestructurados provenientes de diferentes almacenamientos de datos como lo pueden ser bases de datos operativas o lagos de datos. Es un servicio muy fácil de utilizar ya que está basado en lenguaje SQL para realizar las consultas a los datos.



Ilustración 4. Logo de Amazon Redshift.

Redshift cuenta con un sistema de machine learning incluido lo cual permite realizar análisis de datos complejos en cuestión de segundos además de que puede ser conectado con aplicaciones de visualización como lo es Power BI para el uso de usuarios finales. En el apartado *Descripción de cada componente de la solución* de este capítulo se explicará detalladamente la utilidad de Redshift en este proyecto.



Ilustración 5. Logo de Talend Studio.

Talend: Talend es una aplicación de escritorio cuyo principal objetivo es lograr la unificación de los datos, proporcionando para ello funcionalidades como la rápida integración, transformación y mapeo de datos contando además con controles automatizados lo que permite garantizar datos confiables en cada fase. En este proyecto, esta herramienta se ha utilizado como herramienta ETL (extracción, transformación y carga) para llevar los datos resultados de operaciones transaccionales hasta un datawarehouse diseñado con el fin de

que sea útil en la toma de decisiones con respecto a las necesidades de inventario que la empresa presenta. En el apartado *Descripción de cada componente de la solución* de este capítulo se detallarán las fases en las cuales Talend ha sido de utilidad en este proyecto.

Power BI: Es una herramienta de visualización de datos diseñada principalmente para que usuarios finales puedan realizar análisis de datos y de acuerdo a ello tomar las decisiones apropiadas para la organización. Power BI cuenta con una aplicación de escritorio, en la cual se pueden diseñar las visualizaciones de informes de acuerdo a los requerimientos de usuario final. Para diseñar estas visualizaciones Power BI cuenta con soporte para aceptar datos provenientes desde una variedad de

fuentes incluyendo archivos (Excel, XML, csv, entre otros), bases de datos transaccionales como MySql y SQLServer, así como bases de datos en la nube por ejemplo Amazon Redshift. En este proyecto Power BI se ha utilizado para extraer datos contenidos en Amazon Redshift



Ilustración 6. Logo de Power BI.

y con respecto a ellos crear visualizaciones para usuario final. En el apartado *Descripción de cada componente de la solución* de este capítulo se explicará con más detalle cuál es el conjunto de datos utilizado y las visualizaciones creadas.

D. DIAGRAMA ARQUITECTÓNICO DE LA SOLUCIÓN

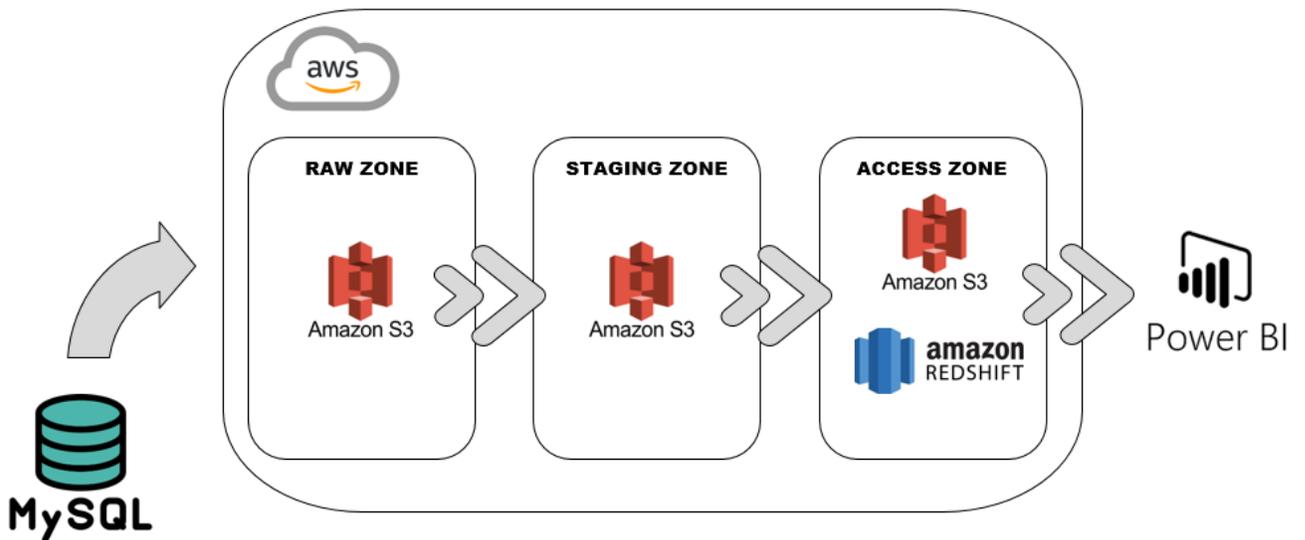


Ilustración 7. Diagrama de la arquitectura de la solución

E. DESCRIPCIÓN DE CADA COMPONENTE DE LA SOLUCIÓN

La solución presentada en el diagrama del apartado anterior cuenta principalmente con tres componentes: La fuente de datos en MySQL, la arquitectura de AWS (la cual contiene tres zonas o fases por las que atraviesan los datos: raw zone o zona de datos crudos, staging zone o zona de procesamiento y access zone o zona de presentación) y la herramienta de visualización Power BI para presentar los resultados. A continuación se detallan cada uno de los componentes mencionados.

Fuente de datos MySQL

Como fuente de datos se cuenta con la base de datos transaccional perteneciente al sistema de inventario PartKeepr que utiliza la empresa El Salvador Courier para almacenar los datos de las transacciones realizadas diariamente sobre el inventario de productos. Esta base de datos tiene por nombre *partkeepr* y utiliza el motor de bases de datos MySQL. La base de datos transaccional del sistema Partkeepr cuenta originalmente con la cantidad de 50 tablas, de las cuales se muestra en el anexo A un dataset de 22 tablas, las cuales fueron pre-seleccionadas como relevantes para realizar el presente trabajo.

Amazon Web Services (AWS)

Amazon Web Services ha sido utilizado en este trabajo como el lugar donde los datos son almacenados primeramente como datos crudos provenientes de la base de datos transaccional *partkeepr* y finalmente como datos procesados listos para realizar visualizaciones o informes en la herramienta Power BI. Para lograr este propósito los datos han sido manejados a través de las tres fases mencionadas al inicio de este apartado y que se explican a continuación.

- **Raw zone:** En esta fase los datos son obtenidos desde la base de datos transaccional *partkeepr* y cargados a la nube de AWS. Los datos en esta fase no sufren alteración alguna en su estructura. Con ayuda de la herramienta Talend Studio se genera una consulta por cada tabla para extraer su contenido y cargar un archivo csv con los datos correspondientes a la nube de AWS. Dentro de AWS se ha creado un bucket de S3 con el nombre *ceid-group8-partkeepr* y una subcarpeta llamada *raw* donde estos archivos csv son almacenados y listos para continuar con la siguiente fase llamada staging o procesamiento.

- **Staging zone:** En esta fase los datos son extraídos de la subcarpeta *raw* del bucket *ceid-group8-partkeepr* de S3 y son transformados de manera que se generen archivos csv con las estructuras apropiadas de acuerdo al diseño de datawarehouse presentado en el apartado *Descripción de la propuesta de solución*. Para cumplir con este objetivo es de crucial importancia la utilización del software Talend Studio, el cual permite realizar la extracción, transformación y carga de datos de forma instantánea. Los archivos csv aquí generados son almacenados en el bucket *ceid-group8-partkeepr* de S3 en la subcarpeta *staging* listos para ser enviados a la zona de presentación.
- **Access zone:** En esta fase los datos son extraídos de la subcarpeta *staging* del bucket *ceid-group8-partkeepr* de S3 y son cargados a la subcarpeta *presentation* del mismo bucket con ayuda del software Talend Studio. Posteriormente, utilizando la misma herramienta, estos datos son extraídos de la subcarpeta *presentation* y cargados en una base de datos creada en Amazon Redshift, la cual lleva por nombre CourierDW y sigue la estructura presentada en el apartado *Descripción de la propuesta de solución*.

Power BI

Esta herramienta se conecta a Amazon Redshift para crear las visualizaciones correspondientes y las cuales son presentadas en el apartado *Análisis de resultados* del capítulo III.

CAPITULO III: ESTRATEGIA DE IMPLEMENTACIÓN DE PROPUESTA DE SOLUCIÓN

A. ESTRATEGIA DE IMPLEMENTACIÓN

La solución presentada en este proyecto se considera un producto nuevo en la empresa El Salvador Courier ya que la empresa no cuenta actualmente con una herramienta que le permita realizar análisis de los datos contenidos en la base de datos transaccional del sistema de inventario Partkeepr utilizado para capturar las operaciones relacionadas con el inventario de las bodegas que la empresa administra.

B. PRESUPUESTO DE IMPLEMENTACIÓN

Costos de implementación de AWS utilizando AWS Pricing Calculator

- **Uso de Amazon S3**

Tiered price for: 500 GB

$500 \text{ GB} \times 0.0230000000 \text{ USD} = 11.50 \text{ USD}$

Costo total de la capa = 11.5000 USD (costo de almacenamiento en S3 Estándar)

$10.000 \text{ Solicitudes PUT para almacenamiento de S3 Standard} \times 0,000005 \text{ USD por solicitud} = 0,05 \text{ USD}$ (costo de solicitudes PUT en S3 Estándar)

$20.000 \text{ Solicitudes GET en un mes} \times 0,0000004 \text{ USD por solicitud} = 0,008 \text{ USD}$ (costo de solicitudes GET en S3 Estándar)

$200 \text{ GB} \times 0,0007 \text{ USD} = 0,14 \text{ USD}$ (costo por las devoluciones de S3 Select)

$200 \text{ GB} \times 0,002 \text{ USD} = 0,40 \text{ USD}$ (costo de escaneados en S3 Select)

$11,50 \text{ USD} + 0,008 \text{ USD} + 0,05 \text{ USD} + 0,14 \text{ USD} + 0,40 \text{ USD} = 12,10 \text{ USD}$ (Total de almacenamiento de S3 Standard, solicitudes de datos, costo de S3 Select)

Costo de S3 Estándar (mensual): 12.10 USD

- **Uso de Amazon Redshift**

$1 \text{ en ejecución} \times 0.25 \text{ USD por hora} \times (60 / 100 \text{ utilizados/mes}) \times 730 \text{ horas en un mes} = 109.5000 \text{ USD}$

Costo de la instancia de Redshift (mensual): 109.50 USD

Servicio	Costo mensual (\$)	Costo anual (\$)
Amazon S3	12.10	145.20
Amazon Redshift	109.50	1,314.00
Power BI	30.00	360.00
Total	151.60	1,819.20

C. ANÁLISIS DE RESULTADOS

Se presentan a continuación dos de las visualizaciones creadas en la herramienta de visualización Power BI.

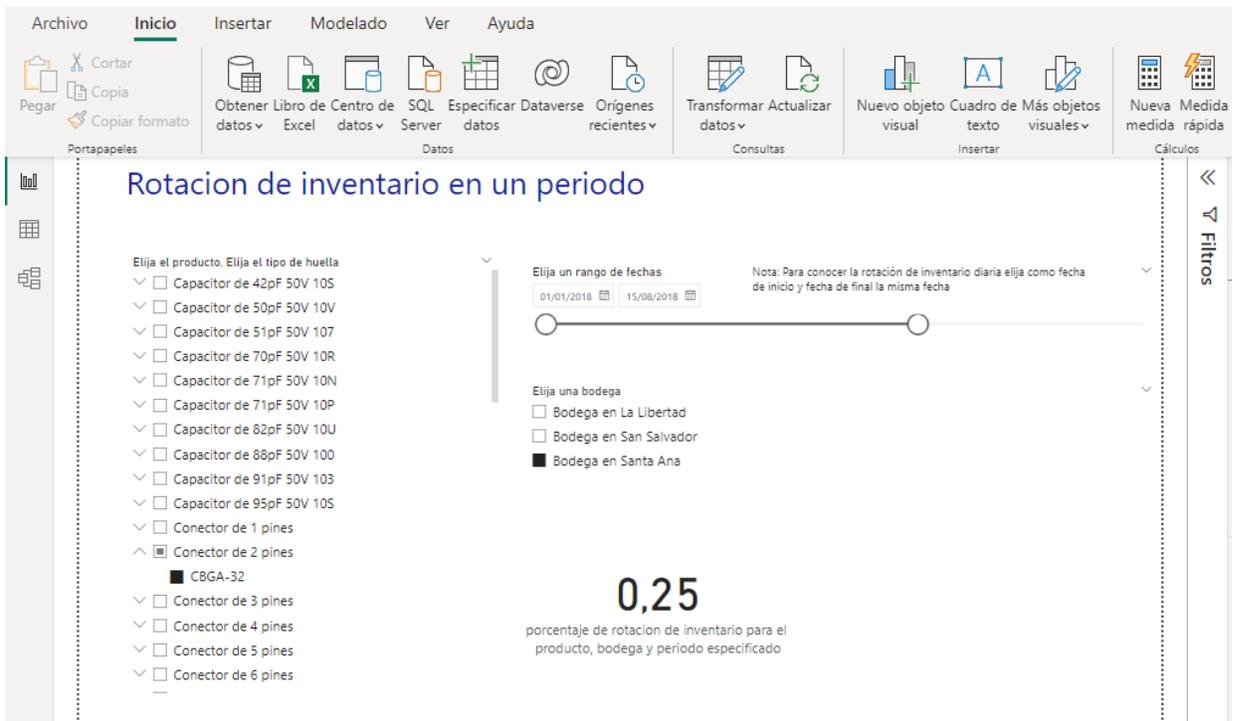


Ilustración 8. Visualización en Power BI de rotación de inventario de un producto en un periodo

Para obtener la rotación de inventario de un producto durante un periodo se ha creado una medida que divide la **suma de cantidades despachadas** del periodo entre el **promedio de cantidades disponibles diariamente** durante ese periodo.

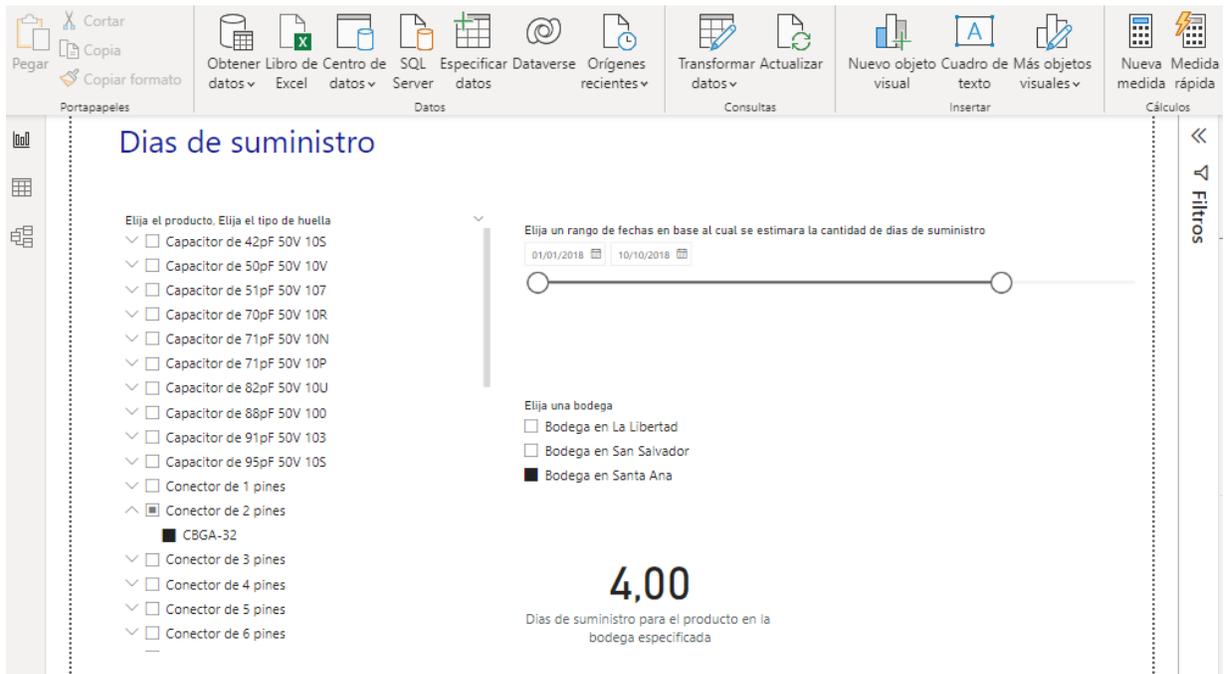


Ilustración 9. Visualización en Power BI de los días de suministro de un producto en una bodega en específico

Para obtener la cantidad de días de suministro de un producto se ha creado una medida que divide la **cantidad disponible que se tiene al final** de un periodo entre el **promedio de cantidades despachadas diariamente** durante ese periodo.

CONCLUSIONES Y RECOMENDACIONES

La implementación de un datawarehouse en conjunto con los servicios en la nube de Amazon Web Services ha traído innegables beneficios a la empresa “El Salvador Courier”, agilizando de manera significativa su proceso para generar reportes, lo que es de gran ayuda para los tomadores de decisiones al momento de ver el rendimiento de la empresa.

Entre los beneficios aportados se tiene la eliminación de la infraestructura física necesaria para implementar este tipo de soluciones, lo que permite ahorrar su costo y mantenimiento gracias a la utilización de servicios en la nube, que son mucho más accesibles en comparación.

La disponibilidad de este tipo de soluciones es de un 99 %, por lo que se puede hacer uso de ella cuando se quiera sin ningún problema. La escalabilidad puede ser ajustada de manera fácil y rápida en cualquier momento a medida que vaya creciendo la cantidad de datos generados, o según cuanta potencia de procesamiento sea requerida.

Aún si ya se parte de un software que es capaz de generar reportes, es recomendable adoptar este tipo de soluciones para estar a la vanguardia y mejorar la productividad aprovechando sus beneficios.

BIBLIOGRAFÍA

Amazon Redshift. La mejor relación precio-calidad para el almacenamiento de datos en la nube. (5 de enero de 2023). Obtenido de <https://aws.amazon.com/es/redshift/>

Amazon S3. Almacenamiento de objetos creado para recuperar cualquier volumen de datos desde cualquier ubicación. (05 de enero de 2023). Obtenido de <https://aws.amazon.com/es/s3/>

Data integration. Bring data together from multiple sources to power all your business decisions. (05 de enero de 2023). Obtenido de <https://www.talend.com/products/integrate-data/>

Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit. The Definitive Guide to Dimensional Modeling.* Wiley.

ANEXOS

Anexo A: Dataset seleccionado de la base de datos Partkeepr

