Proyecto de Práctica profesional

Estudiantil Orientado a la Asignatura

Base de

Datos Avanzada.

Título: Implementación de un Data WareHouse en Power BI en la empresa GeoCuba Oriente-Sur

Nombre: Dariel Enmanuel Cabrera López

Carrera: Ingeniería Informática

Profesor: MSc. Pedro Antonio Suárez Salfrán

Fecha de Realización

Resumen

Las costas y zonas de playas son un importante recurso para el desarrollo humano. Grandes núcleos poblacionales se desarrollan a lo largo de las costas y las playas resultan un recurso importante para el sector turístico y de materiales de la construcción, entre otros.

El cambio climático, en especial el aumento del nivel del mar ha traído consigo el deterioro de las costas y el cambio constante de la línea que la define. Tener herramientas que puedan medir con efectividad este deterioro es importante en la toma de decisiones.

En este trabajo se propone un sistema de almacenamiento diseñado para la empresa Geocuba (Sucursal Oriente Sur) que …

Palabras claves: Sistema informático, toma de decisiones, Geocuba, cambio climático

Introducción

El cambio climático ha sido causado por las actividades antrópicas que durante cientos de años se han desarrollado a nivel mundial, estas han ocasionado un impacto negativo y considerable como el aumento de la temperatura de la Tierra relacionado directamente con la emisión de gases de efecto invernadero (GEI) generados por las actividades productivas, sistemas de transporte con uso de combustibles fósiles, inadecuada gestión de residuos sólidos y en general, por el uso desmesurado de recursos naturales y pocas prácticas de sostenibilidad ambiental. (Medina, s. f.)

Este es una amenaza real para las costas de todo el mundo. El aumento del nivel del mar, la erosión costera y la acidificación de los océanos son algunos de los efectos más graves del cambio climático en las costas

En Cuba la erosión costera es un problema grave. Según en el Segundo Periodo Ordinario de Sesiones de la IX Legislatura de la Asamblea Nacional del 2018 se estimó que e**l 85% de las playas arenosas cubanas presentan indicios de erosión: 23 con erosión intensa y 195 con erosión moderada** (*Cubadebate*, s. f.-a)**. De 505 playas, se han evaluado 247 y se ha comprobado que 208 (84.2 %) presentan indicios de erosión.**(*Cubadebate*, s. f.-b).Para enfrentar estos efectos la isla aprobó en 2017 La Tarea Vida que es un plan del estado cubano para el enfrentamiento del cambio climático. Además de esto el gobierno cubano aprobó el Decreto-Ley 77 de las Costas y su Reglamento en 2023.(*Nueva norma legal por la preservación de las costas cubanas - Juventud Rebelde - Diario de la juventud cubana*, s. f.)

Una de las empresas que está realizando aportes a la tarea Vida es la empresa Geocuba. [GEOCUBA es un grupo empresarial cubano que se dedica a la elaboración y comercialización de información, tecnologías, productos y servicios en diversas esferas](https://www.ecured.cu/GEOCUBA) (*GEOCUBA - EcuRed*, s. f.)

El cálculo transporte de sedimentos en las costas cubanas es la tarea que realizan los especialistas de la empresa Geocuba para investigar, predecir y monitorear el comportamiento de la línea costera y la erosión del mar en zonas de desarrollo de la economía y que no afecte el medio ambiente.

Desarrollo Teórico

En esta sección, se abordarán aspectos teóricos relevantes relacionados con BI, las teorías que la sustentan, los Almacenes de datos ,Datamart, modelos dimensionales, herramientas de extracción y transformación de datos, así como técnicas de análisis de datos. Se profundizará en la importancia de los Datamarts en el contexto empresarial, su estructura y utilidad para la generación de información estratégica. Tres hojas para el desarrollo teórico como mínimo y cinco máximo

**Business Intelligence*(BI)***

Las aplicaciones de Business Intelligence (BI) son herramientas de soporte de decisiones que permiten en tiempo real, acceso interactivo, análisis y manipulación de información crítica para la empresa. Estas aplicaciones proporcionan a los usuarios un mayor entendimiento que les permite identificar las oportunidades y los problemas de los negocios. Los usuarios son capaces de acceder y apalancar una vasta cantidad de información y analizar sus relaciones y entender las tendencias que últimamente están apoyando las decisiones de los negocios. Estas herramientas previenen una potencial pérdida de conocimiento dentro de la empresa que resulta de una acumulación masiva reinformación que no es fácil de leer o de usar.  (CherryTree & Co., 2000)

**Data Warehouse (DW)**

Un almacén de datos es un sistema de gestión de datos diseñado para respaldar las actividades de inteligencia empresarial (BI), en particular el análisis. Centraliza y consolida grandes cantidades de datos de múltiples fuentes, lo que permite a las organizaciones obtener información valiosa para la toma de decisiones. Los datos dentro de un almacén de datos suelen ser históricos y provienen de varias fuentes, como registros de aplicaciones y aplicaciones de transacciones. Sirve como la "única fuente de verdad" de una organización mediante la creación de un registro histórico que es invaluable para los científicos de datos y los analistas de negocios. Un almacén de datos típico incluye una base de datos relacional para el almacenamiento de datos, una solución ELT (extracción, carga y transformación) para la preparación de datos, el análisis estadístico, la generación de informes y las capacidades de minería de datos, herramientas de análisis de clientes para la visualización de datos y aplicaciones analíticas más sofisticadas que aplican algoritmos de ciencia de datos e IA. (*What Is a Data Warehouse? | Oracle*, s. f.)

Según (Alarcón Crespo, 2022) definen al Data Warehouse como una base de datos corporativos que se caracteriza por integrar y depurar la información para luego procesarla, permitiendo el análisis de la información desde distintos puntos de vista. Además (Alarcón Crespo, 2022), expresan que la creación de un DW representa en la mayoría de las ocasiones el primer paso, desde el 10 punto de vista técnico, para implantar una solución completa y fiable de Business Intelligence.

**Modelo Dimensional**

El modelo Dimensional, es el diseño físico y lógico que transformará las antiguas Fuentes de datos en estructuras finales del DW a través.

De una técnica que busca la presentación de los datos en un marco de trabajo estándar que es intuitivo y permite un acceso al alto desempeño. Cada modelo dimensional está formado de una tabla tiene una llave compuesta llamada tabla de hechos y conjunto de tablas más pequeñas llamadas dimensiones. Cada tabla dimensión tiene una llave primaria simple, que corresponde exactamente a una de las partes de la llave compuesta en la tabla de hechos. Esta estructura característica es usualmente llamada esquema de estrella. (Vargas-Machuca et al., s. f.)

**Características**

El modelado dimensional es una técnica de diseño lógico, esta técnica nos proveerá de una estructura de los datos que tengan las siguientes características:

• Un alto rendimiento en la consulta de datos

• Intuitivos para los usuarios del negocio. Los modelos dimensionales son llamados modelos o esquemas estrellas, estos modelos son guardados en estructuras OLAP conocidas como Cubos OLAP. 19

Proceso iterativo para crear un modelo dimensional:

1. Elegir el proceso del negocio: Este proceso depende del análisis de requerimientos del negocio.

2. Establecer el nivel de granularidad: El nivel de granularidad depende del nivel de detalle que se quiera obtener, se recomienda buscar el nivel de detalle más profundo que permitan los datos, tomando en cuenta lo que se necesite según diga el análisis de requerimientos del negocio.

3. Elegir dimensiones: Las dimensiones tienen generalmente atributos textuales que nos brindaran el contexto para el nivel de granularidad escogido. Para identificar las dimensiones se deben analizar sus atributos a fin de encontrar atributos candidatos a ser encabezados de informes, cubos o cualquier tipo de visualización unidimensional o multidimensional.

4. Identificar medidas y tablas de hechos: Las medidas o métricas son aquellos valores que surgen en los procesos de negocios, una medida o métrica es un atributo que se desea analizar, sumarizando o agrupando los datos. Estas medidas o métricas estarán almacenadas en las tablas de hechos, estas tablas están relacionadas con las dimensiones que proveen del contexto de las medidas. (Godoy Gutierrez et al., 2022)

**Power BI**

Esta es una herramienta que admite combinar diversas bases de datos, para después crear sistemas que analicen la información que contiene. Entonces, la información se puede presentar de manera que los formatos sean fáciles de entender y emplear. Power BI es una solución de investigación que le permite ver y compartir datos, conectarse a múltiples bases de datos y generar potentes componentes e informes. Además, proporcionamos servicios desde una aplicación de escritorio de Windows llamada Power BI Desktop, un servicio en internet, software como servicio (SaaS), llamado Servicio Power BI, o una aplicación móvil Power BI disponible en teléfonos móviles y tablets. Con BI Desktop, puede crear plantillas de datos, crear vistas e informes, compartir archivos de informes o incrustarlos en aplicaciones y sitios web(Guerra & Molina, 2021).

**Importancia del Data Warehouse**

La importancia de un Data Warehouse radica en su papel fundamental como base para el análisis de datos en Business Intelligence (BI). Un Data Warehouse es un sistema especializado que sirve como un repositorio centralizado, confiable e integrado para el almacenamiento de datos históricos y consolidados.

**Fuente Única de Verdad**: Los Data Warehouses eliminan la necesidad de navegar por datos dispersos en diferentes sistemas, actuando como una única fuente de verdad. Esto asegura la consistencia y precisión de los datos, reduciendo el riesgo de errores o conclusiones engañosas y permitiendo a BI ofrecer insights confiables y accionables.

**Mejora de la Calidad de los Datos**: El proceso de transformación de datos permite limpiar, estandarizar e integrar información de diversas fuentes, resultando en datos de alta calidad adecuados para análisis confiables. Además, la naturaleza no volátil de los Data Warehouses preserva la integridad de los datos históricos para análisis a largo plazo.

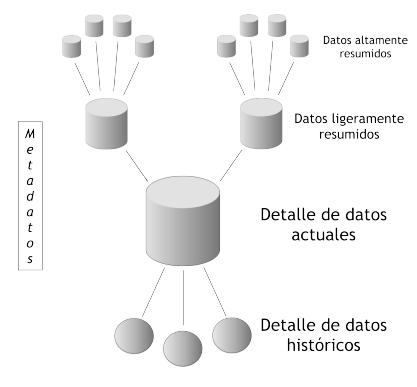
**Análisis más Rápido**: Diseñados para utilizar procesamiento paralelo y almacenamiento de datos en columnas, los Data Warehouses permiten una recuperación y análisis más rápido de grandes conjuntos de datos. Esto permite dedicar menos tiempo a buscar datos relevantes y más tiempo a extraer insights valiosos de la información.

**Consultas Complejas y Exploración**: La arquitectura estructurada del Data Warehouse facilita la ejecución eficiente de consultas complejas y la exploración profunda de los datos. Esto permite profundizar en los datos, descubrir patrones ocultos y explorar diversas dimensiones, alimentando la toma de decisiones estratégicas con insights completos que responden a preguntas complejas del negocio

**Organización por Temas**: A diferencia de las bases de datos operativas, que se centran en operaciones diarias y transacciones, los Data Warehouses organizan los datos alrededor de temas comerciales como ventas, marketing, servicio al cliente, etc. Este enfoque temático simplifica el análisis al permitir enfocarse en áreas de interés específicas, facilitando el descubrimiento de tendencias y patrones a lo largo del tiempo. (*What is the Role of Data Warehouse in Business Intelligence? | Airbyte*, s. f.)

Un Data Warehouse es esencial para el análisis de datos en BI, proporcionando una plataforma centralizada, confiable e integrada para el almacenamiento y análisis de datos. Su capacidad para consolidar datos de múltiples fuentes, mejorar la calidad de los datos, permitir análisis rápidos y facilitar consultas complejas y exploraciones profundas, lo convierte en un componente crítico para la toma de decisiones informadas y el crecimiento estratégico de las organizaciones.

**Estructura del Data Warehouse**

****

* **Detalle de datos actuales**: son aquellos que reflejan las ocurrencias más recientes. Generalmente se almacenan en disco, aunque su administración sea costosa y compleja, con el fin de conseguir que el acceso a la información sea sencillo y veloz, ya que son bastante voluminosos. Su gran tamaño se debe a que los datos residentes poseen el más bajo nivel de granularidad, o sea, se almacenan a nivel de detalle. Por ejemplo, aquí es donde se guardaría el detalle de una venta realizada en tal fecha.
* **Detalle de datos históricos:** representan aquellos datos antiguos, que no son frecuentemente consultados. También se almacenan a nivel de detalle, normalmente sobre alguna forma de almacenamiento externa, ya que son muy pesados y en adición a esto, no son requeridos con mucha periodicidad. Este tipo de datos son consistentes con los de Detalle de datos actuales. Por ejemplo, en este nivel, al igual que en el anterior, se encontraría el detalle de una venta realizada en tal fecha, pero con la particularidad de que el día en que se registró la venta debe ser lo suficientemente antigua, para que se considere como histórica
* **Datos ligeramente resumidos**: son los que provienen desde un bajo nivel de detalle y sumarizan o agrupan los datos bajo algún criterio o condición de análisis. Habitualmente son almacenados en disco. Por ejemplo, en este caso se almacenaría la sumarización del detalle de las ventas realizadas en cada mes
* **Datos altamente resumidos:** son aquellos que compactan aún más a los datos ligeramente resumidos. Se guardan en disco y son muy fáciles de acceder. Por ejemplo, aquí se encontraría la sumarización de las ventas realizadas en cada año.
* **Metadatos:** representan la información acerca de los datos. De muchas maneras se sitúa en una dimensión diferente al de otros datos del DW, ya que su contenido no es tomado directamente desde el ambiente operacional.(Bernabeu, 2010)

Utilidad

Un Data Warehouse es una herramienta esencial para las empresas que manejan grandes volúmenes de datos provenientes de múltiples fuentes. Su utilidad radica en varias áreas clave:

* **Organización y Acceso Facilitado**: Permite almacenar datos de manera organizada y estructurada, facilitando su acceso y análisis. Esto es especialmente útil para empresas que buscan mejorar la eficiencia en la toma de decisiones empresariales, ya que los datos están disponibles de manera categorizada y fácil de entender
* **Soporte a la Toma de Decisiones**: Al proporcionar una visión consolidada de los datos de la empresa, los Data Warehouses ayudan a los ejecutivos de negocios a tomar decisiones estratégicas basadas en información precisa y actualizada. Esto incluye la creación de informes, diagramas y gráficos que ilustran tendencias y patrones en los datos
* **Prevención de Dark Data**: Evita la acumulación de datos no utilizados (Dark Data) al asegurar que todos los datos sean almacenados de manera efectiva y sea fácil recuperarlos cuando sea necesario. Esto ayuda a las empresas a maximizar el valor de sus datos y evitar el desperdicio de recursos en almacenamiento innecesario
* **Alimentación con Datos Confiables**: Los Data Warehouses pueden ser nutridos con datos confiables y verídicos a través de técnicas de investigación de mercado. Esto asegura que las decisiones tomadas basadas en los datos sean informadas y basadas en evidencia sólida [**1**](https://www.rochiconsulting.com/blog/data-warehouse/).
* **Adaptabilidad a Diferentes Industrias**: La utilidad de un Data Warehouse no está limitada a un tipo específico de industria. Puede ser adaptado a las necesidades de cualquier tipo de empresa, lo que lo convierte en una herramienta versátil para el manejo de datos
* **Seguridad y Fidelidad en el Almacenamiento**: Los datos almacenados en un Data Warehouse deben ser seguros, fiables, fácilmente recuperables y administrables. Esto garantiza que las empresas puedan confiar en la integridad y disponibilidad de sus datos para análisis y decisiones. (*Utilidad del Data Warehouse para las Empresas - rochiconsulting*, s. f.)

Metodologia

La metodología a seguir para la realización de este proyecto comprenderá las siguientes etapas:

1. Investigación y análisis del contexto organizacional y de los requerimientos de información.
2. Diseño del modelo dimensional y definición de dimensión y métricas clave.
3. Extracción, transformación y carga de los datos en el Datamart.
4. Implementación y prueba del Datamart con herramientas específicas.
5. Desarrollo de consultas y análisis para la obtención de información relevante.
6. Documentación detallada del proceso y resultados obtenidos.

**Caso de Estudio**

La empresa Geocuba Oriente Sur desarrolla cálculos del transporte litoral a nivel provincial y nacional. Su objetivo principal es determinar cuál el comportamiento de los sedimentos en una zona de estudio.

La empresa necesita mejorar su eficiencia en la toma de decisiones y contar con información detallada a tal fin.

**Paso 1: Análisis de Requerimientos**

**Proceso** Cálculo:

Se desea conocer como es el volumen de sedimentos (Q) con valores que poco varían en el tiempo en una zona de estudio en un periodo de tiempo determinado.

Volumen de sedimentos (Q) en una **zona** en un **tiempo** determinado

Se desea conocer cómo se ajusta la variable empírica (K) para esa zona en una zona de estudio en un periodo determinado

Valor de la variable empírica (K) en una zona estudio en un periodo de tiempo determinado

**Identificar indicadores y perspectivas de análisis**

Volumen de sedimentos (Q) en una **zona** en un **tiempo** determinado

**Valor de la variable empírica (K)** en una **zona** estudio en un periodo de **tiempo** determinado

**Modelo Conceptual**

Volumen de

Sedimentos

Zona

Calculo

Valor de variable K

Tiempo

**Paso 2 Análisis de los de los OLTP (On Line Transaction Processing)**

OLTP representa toda aquella información transaccional que genera la empresa en su accionar diario, además de las fuentes externas con las que puede llegar a disponer.

**Paso 3: Proceso ETL**

Construcción

Durante esta etapa, se llevará a cabo la construcción del Datamart siguiendo el modelo dimensional diseñado, incorporando los datos necesarios y garantizando la calidad y coherencia de la información almacenada. Emplear PDI, Schema Workbench y Report Designer.

Conclusiones

Tras la implementación del Datamart y la realización de análisis pertinentes, se presentarán las conclusiones obtenidas, destacando la utilidad y eficacia del Datamart en la generación de información estratégica para la organización. Una página para las conclusiones

Bibliografía

Alarcón Crespo, L. Á. (2022). Modelo de un Sistema de Información aplicando Data Warehouse para el control del Tráfico en el sector norte de la ciudad de Guayaquil.

Bernabeu, R. D. (2010). Hefesto–Data warehousing: Investigación y sistematización de conceptos-Hefesto: Metodología para la construcción de un data warehouse. Córdova, Argentina.

Cubadebate. (s. f.-a). Recuperado 7 de enero de 2024, de http://www.cubadebate.cu/noticias/2018/12/16/tarea-vida-a-debate-en-la-asamblea-nacional-cuba-ante-las-realidades-del-cambio-climatico/

Cubadebate. (s. f.-b). Recuperado 7 de enero de 2024, de http://www.cubadebate.cu/noticias/2023/07/18/tarea-vida-en-la-mira-de-los-diputados-hay-una-transicion-en-el-clima-de-cuba/

GEOCUBA - EcuRed. (s. f.). Recuperado 9 de enero de 2024, de https://www.ecured.cu/GEOCUBA

Godoy Gutierrez, H. I., Miranda Miranda, N. E., & Presa Mariona, J. R. (2022). Desarrollo e implementación de un modelo dimensional para el proceso de negocio de preguntas y respuestas de la plataforma Stack Overflow.

Guerra, L. G. O., & Molina, M. A. C. (2021). Propuesta de un Data Warehouse para el análisis de la circulación bibliográfica en la biblioteca de la Universidad Nacional de Educación. Dominio de las Ciencias, 7(3), 615-632.

Medina, N. M. C. (s. f.). EL CAMBIO CLIMÁTICO Y SUS EFECTOS SOBRE LA SALUD HUMANA.

Nueva norma legal por la preservación de las costas cubanas—Juventud Rebelde—Diario de la juventud cubana. (s. f.). Recuperado 15 de abril de 2024, de https://www.juventudrebelde.cu/ciencia-tecnica/2023-12-07/acertada-interrelacion-entre-las-entidades-y-rigor-en-el-control-contribuiran-al-exito-de-la-nueva-norma-legal-relativa-a-las-costas-cubanas

Utilidad del Data Warehouse para las Empresas—Rochiconsulting. (s. f.). Recuperado 9 de mayo de 2024, de https://www.rochiconsulting.com/blog/data-warehouse/

Vargas-Machuca, B. A. B., Burgos, A. A. T., Duque, L., & Reyes, P. (s. f.). DESARROLLO DE UNA APLICACIÓN DE BUSINESS INTELLIGENCE (BI) PARA LA EMPRESA EMPAQPLAST.

What Is a Data Warehouse? | Oracle. (s. f.). Recuperado 17 de abril de 2024, de https://www.oracle.com/database/what-is-a-data-warehouse/

What is the Role of Data Warehouse in Business Intelligence? | Airbyte. (s. f.). Recuperado 18 de abril de 2024, de https://airbyte.com/data-engineering-resources/business-intelligence-data-warehouse