|  |  |
| --- | --- |
| PROGRAMA DE FORMACIÓN | Tecnólogo en implementación y gestión de bases de datos |

|  |  |  |  |
| --- | --- | --- | --- |
| COMPETENCIA | 210602033 - Alimentar la base de datos de acuerdo con procedimientos técnicos. | RESULTADOS DE APRENDIZAJE | 210602033-01 Definir la arquitectura de la bodega de datos según características del proyecto a desarrollar y los requerimientos técnicos. |

|  |  |
| --- | --- |
| NÚMERO DEL COMPONENTE FORMATIVO | CF5 |
| NOMBRE DEL COMPONENTE FORMATIVO | Bodegas de datos |
| BREVE DESCRIPCIÓN | Actualmente, todas las organizaciones generan datos. Los sistemas informáticos crecen continuamente y la disponibilidad 24/7 de la información es clave en la toma de decisiones. Una información dinámica, oportuna y accesible agiliza procesos y produce valor agregado y elecciones acertadas. Así, construir una bodega de datos para su gestión veloz, voluminosa y variada brindará productividad y mejores beneficios al negocio. |
| PALABRAS CLAVE | Arquitectura, bodegas de datos, fuentes de datos, *big data*, *data marts* |

|  |  |
| --- | --- |
| ÁREA OCUPACIONAL | Ciencias naturales, aplicadas y relacionadas |
| IDIOMA | Español |

# **Tabla de contenidos**

**Introducción**

**1. Introducción a la bodega de datos**

1.1. Definición

1.2. Elementos de una bodega de datos

1.3. Implementación de bodega de datos

**2. Diseño de bodega de datos**

2.1. Sistema operativo y red

2.2. OLAP

2.3. Característica y entorno técnico del negocio

2.4. Requerimientos del usuario

**3. Arquitectura de una bodega de datos**

3.1. Arquitectura de una capa

3.2. Arquitectura de dos capas

3.3. Arquitectura de tres capas

3.4. *Data marts*

**4. Implementación de una bodega de datos**

4.1. Interfases

4.2. Usuarios

4.3. Fuentes de datos

**Introducción**

|  |
| --- |
| Cuadro de texto |
| Apreciado aprendiz, reciba una cordial bienvenida al componente formativo Bodegas de datos. Aquí usted contará con orientación permanente para el desarrollo de este programa. Además, tendrá la oportunidad de aprender y aplicar las técnicas para diseñar y construir una bodega de datos, saber qué tipo de problemas puede resolver con estos procedimientos y así aplicarlos en un caso específico.  ¡Muchos éxitos en este proceso de aprendizaje! |

**Guion de video introductorio**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tipo de recurso** | Video motion | | | |
| **NOTA** | **La totalidad del texto locutado para el video no debe superar las 500 palabras aproximadamente** | | | |
| **Título** | Bodegas de datos | | | |
| **Escena** | **Imagen** | **Sonido** | **Narración** | **Texto** |
| **1** | **Fuente:** <https://shutr.bz/3EKioYm> | Música de fondo | Internet y las nuevas tecnologías han facilitado el acceso y el almacenamiento excesivo de información, con una gran velocidad, volumen y variedad. Al mismo tiempo, estas herramientas se han convertido en un activo muy importante, lo cual ha hecho que las empresas sean cada vez más conscientes de la importancia que tienen los datos. De hecho, algunas organizaciones no han modelado su sistema de información de acuerdo con la necesidad que tienen y cuentan con escasa experiencia en el manejo de datos. Esto, por supuesto, les dificulta la tarea de consolidarlos, por la falta de estándares y por contar con bases de datos dispersas. En consecuencia, las empresas son susceptibles de sufrir inconvenientes y de enfrentar múltiples problemas al momento de realizar una consulta intensiva de información; igualmente, pueden tener dificultades en el nivel de servicio del resto del sistema, lo que amenaza su rentabilidad. Por tanto, hay organizaciones que ponen en riesgo toda su información por no contar con un control minucioso de su gestión y almacenamiento. | Gestión y almacenamiento de la información. |
| **2** | **Fuente:** <https://shutr.bz/3GwzSZr> | Música de fondo | Para que una empresa sea más competitiva, tome mejores decisiones y diseñe estrategias acertadas, necesita acceder de manera rápida y fácil a la información, en medio de un contexto dinámico, competitivo y con alto volumen de crecimiento de los datos, que recibe a diario de diferentes canales. Así que es imprescindible desarrollar una infraestructura informática que centralice o cargue todo ese volumen de datos en un esquema unificado, que permita en paralelo manejar la información de manera permanente, oportuna, dinámica y accesible. Se trata de una solución informática que consolide todos los datos obtenidos de diferentes fuentes, ya sea de bases de datos, archivos planos o de otros sistemas del negocio; es decir, que toda la información suministrada esté bajo estándares que permitan su fácil integración y consolidación, que optimicen las consultas y puedan transformar los datos fácilmente, con el fin de apoyar y acelerar la toma de decisiones estratégicas. | Tener estrategias acertadas para acceder de manera rápida y fácil a la información almacenada. |
| **3** | **Fuente:** <https://shutr.bz/3AxMPOR> | Música de fondo | Gracias a las tecnologías de la información, muchas entidades gubernamentales y empresas de servicios públicos, de entretenimiento, de petróleo y gas, editoriales, laboratorios farmacéuticos, droguerías, etc., buscan incrementar su productividad y ventaja competitiva. Esto ha incentivado la automatización de los procesos y ha facilitado los procedimientos de las organizaciones mediante herramientas y estructuras que integran y almacenan datos estratégicos, tácticos y operativos, que son de gran utilidad para que las empresas puedan ser consultadas con mayor velocidad y precisión. Actualmente, las aplicaciones, los teléfonos celulares, las páginas web, las agendas electrónicas, etc., utilizan toda una tecnología de bases de datos para asegurar la integridad de estos y trabajar con altos volúmenes en tiempo real y de manera organizada. | Se busca incrementar la productividad incentivando la automatización de los procesos. |
| **4** | **Fuente:** <https://shutr.bz/3XczBAZ> | Música de fondo | En este componente formativo encontrará los elementos, modelos y componentes de la arquitectura de una bodega de datos; conocerá técnicas para el diseño e implementación de las mismas, siguiendo buenas prácticas de la industria, y sabrá qué pasos seguir con la menor cantidad posible de errores y problemas para lograr una instalación exitosa y con todos los estándares, que aporten valor agregado a las empresas. | La implantación de buenas prácticas en la industria y su correspondiente almacenamiento de datos. |
| **Nombre del archivo** | 228130\_v1 | | |  |

# **1. Introducción a la bodega de datos**

## 1.1. Definición

|  |
| --- |
| Cuadro de texto |
| ¿Qué es una bodega de datos?  También llamado *data warehouse* (DW), es un sistema de almacenamiento digital de datos que maneja grandes cantidades de información, de diferentes fuentes, con el objetivo de alimentar los procesos de inteligencia de negocios (BI, por sus siglas en inglés), informes, consultas y analíticas (SAP, s. f.). Del mismo modo, el DW se encarga de dar soporte a los diferentes requerimientos de manera que las empresas puedan convertir sus datos en información valiosa y tomar mejores decisiones a partir de estos. En suma, un DW guarda los datos tanto históricos como actuales en una sola ubicación y son la única fuente confiable de información para una organización (SAP, s. f.). |

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 1 |
| **Introducción** | Una bodega de datos es una tecnología que acumula información diversa y de manera central de una empresa y que facilita la integración y procesamiento de todos los registros. ElDW brinda un valor agregado de gran utilidad para generar consultas e informes y optimiza la toma de decisiones de una organización. Asimismo, este almacén electrónico registra todas las actividades o la vida de una empresa de manera fácil y rápida para su análisis, lo cual contribuye a la toma de decisiones. |
| **Fuente:** <https://shutr.bz/3gk0UIZ>  **Imagen:** 228130\_i1 | |
| ¿Cuál es su finalidad?:   * Almacenar datos de calidad que sean útiles para su uso en los análisis de la empresa. * Mantener un registro completo de las actividades de la empresa. * Ayudar al funcionamiento de las aplicaciones que permiten la toma de decisiones. | |
| Los siguientes son los tipos de datos en un DW:  Datos de funcionamiento de la empresa: se producen por aplicaciones corporativas, como las que la organización utiliza para satisfacer los pedidos de productos de los clientes o para gestionar las transacciones financieras. Son la materia prima para un almacén de datos. | |
| Integración de los datos del negocio: se construye para mejorar la calidad y sincronizar dos o más aplicaciones corporativas, como una lista maestra de clientes. Los datos se aprovechan para integrar aplicaciones que no fueron diseñadas para funcionar entre sí. | |
| Datos de seguimiento del negocio: presentados a los usuarios finales para la elaboración de informes y apoyo a la toma de decisiones, como su cuadro de mando financiero. Los datos se limpian para permitir a los usuarios comprender mejor el progreso y evaluar las relaciones causa-efecto de los datos. | |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Slider pasos | |
| **Introducción** | En la historia de la bodega de datos, este concepto surge a finales de los años ochenta, como respuesta a la necesidad de facilitar la consolidación de información en los sistemas de apoyo a la decisión o *decision support systems* (DSS), con el trabajo de los investigadores de IBM, Barry Devlin y Paul Murphy (Biscobing, 2021). Aunque fue William H. Inmon, conocido como el padre del*data warehousing*, quien acuñó el término *data warehouse,* desde las décadas de los setenta y ochenta los datos comenzaron a expandirse y las organizaciones necesitaban una forma fácil de almacenar y acceder a dicha información (SAS, s. f.). | |
| **Slide 1** | En 1992, William H. Inmon publicó su libro *Building the Data Warehouse* (*Construcción del almacén de datos*), declarado como principio fundamental de la tecnología de los almacenes de datos (SAS, s. f.). En este último concepto Inmon utiliza un enfoque “descendente”, “donde primero se establece un repositorio centralizado, y luego se crean depósitos de datos – que contienen subconjuntos de datos específicos – dentro de ese repositorio” (SAS, s. f., párr. 2).  Por otro lado, a mediados de los noventa, Ralph Kimball, otro experto en tecnología, publicó *The Data Warehouse Toolkit* (*Herramientas para almacenes de datos*). Kimball empleó una táctica un poco diferente sobre los *data warehousing*, con un enfoque “ascendente”, “donde se integran depósitos de datos individuales para crear un data warehouse” (SAS, s. f., párr. 3). | **Fuente:** <https://alchetron.com/Bill-Inmon>    **Fuente:** <https://www.kimballgroup.com/about-kimball-group>  **Imagen:** 228130\_i2 |
| **Slide 2** | Cuando se trata del [diseño de bodegas de datos (DWH, por sus siglas en inglés)](https://www.astera.com/es/tipo/blog/arquitectura-del-almac%C3%A9n-de-datos/), existen dos de los más usados enfoques de almacenamiento: el de Inmon y el de Kimball. Durante años se ha debatido sobre qué metodología es mejor y más efectiva para las empresas; sin embargo, ambas tienen sus ventajas y desventajas (Naeem, 2020). | **Fuente:** <https://shutr.bz/3AyQ8We>  **Imagen:** 228130\_i3 |
| **Slide 3** | 1. Metodología Inmon: cuenta “con un esquema que parte de la construcción del modelo de datos corporativos, elaborado al más alto nivel de abstracción, para luego derivar la estructura del modelo de datos, para el diseño de la bodega” (Bahamón, 2003, p. 15). | **Figura 1**  *Referente de la metodología de Inmon*    **Imagen:** 228130\_i4 |
| **Slide 4** | 2. Metodología Kimball: cuenta “con un esquema centrado en la identificación de los procesos de la empresa, como elemento clave para la definición de la estructura de variables y dimensiones” (Bahamón, 2003, p. 15). | **Figura 2**  *Referente de la metodología de Kimball*    **Imagen:** 228130\_i5 |
| **Slide 5** | El almacén de datos proporciona la plataforma para implementar, gestionar y entregar estos activos de datos clave. Por tanto, es el proceso de creación de una solución a la gestión de los datos por medio de una arquitectura que permita el procesamiento analítico y de la información, a pesar de las barreras de plataforma, aplicación, organización, entre otras. | **Fuente:** <https://shutr.bz/3EoEQVt>  **Imagen:** 228130\_i6 |
| **Slide 6** | Tipos de almacenes de datos (DWH)  Existen tres tipos principales de DWH, que se utilizan especialmente en sistemas empresariales:   1. Almacén de datos empresariales (EDW): es centralizado y suministra un enfoque integral para establecer y presentar datos. 2. Almacén de datos operativos (ODS): es adecuado y puede respaldar los informes de una empresa cuando el procesamiento de transacciones en línea (OLTP, por sus siglas en inglés) o el DWH no lo puedan hacer. 3. Mercado de datos: diseñado para datos en los departamentos de ventas, finanzas y cadena de suministro. | **Fuente:** <https://shutr.bz/3tHq1IR>  **Imagen:** 228130\_i7 |
| **Slide 7** | Las bodegas de datos se basan en un modelo multidimensional. ¿Qué es una base de datos multidimensional?  Es una matriz de dimensiones que permite guardar y recuperar eficientemente un gran volumen de datos, relacionados, vistos y analizados desde diferentes perspectivas. Ofrecen, además, una mejor comprensión y consulta de la información por medio de análisis complejos. | **Fuente:** <https://bit.ly/3EvYWxk>  **Imagen:** 228130\_i8 |
| **Slide 8** | ¿Qué es un modelado dimensional?  Es una matriz que contiene todos los datos de una forma sencilla, estándar y accesible. Se compone de una tabla de hechos y varias tablas de dimensión. | **Figura 3**  *Referente de un modelado dimensional*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | *Tabla de dimensión* |  | *Tabla de hechos* |  | *Tabla de dimensión* | | **CLIENTE** |  |  |  | **PRODUCTO** | | Id\_Cliente |  | **VENTAS** |  | Id\_Producto | | Nombre\_Cliente |  | ld\_Cliente |  | Nombre\_Producto | | Teléfono\_Cliente |  | ld\_Producto |  | Descripción\_Producto | | Dirección\_Cliente |  | Cantidad\_Producto |  |  |   **Imagen:** 228130\_i9 |

## 1.2. Elementos de una bodega de datos

|  |  |
| --- | --- |
| **Tipo de recurso** | Infografía estática |
| **Texto introductorio** | Un almacén de datos se compone de muchos elementos diferentes y cada uno se implementa de varias maneras. Estos incluyen la información ilustrada por la siguiente infografía: |
| **Imagen**    Por favor elaborar una infografía parecida o igual a la que se presenta en la imagen, se adicionan los archivos de la elaboración de la infografía hecha en canvas, los archivos están en jpg y pdf, y los encuentran en la carpeta con la siguiente denominación CF5\_228130\_Anexo1. | |
| **Código de la imagen** | 228130\_i10 |

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 1 |
| **Introducción** | Entre los elementos que existen en la arquitectura de una bodega de datos, se presentan a continuación seis de ellos: |
| **Figura 4**  *Arquitectura de una bodega de datos*    **Imagen:** 228130\_i11 | |
| Datos internos: es la información almacenada directamente por la empresa; por lo tanto, difícilmente, la competencia la tendrá. | |
| Datos externos: es la información que no depende de la empresa, pero que ayuda de forma valiosa. Se puede obtener ya sea de forma cedida o comprada. | |
| ETL: conocido así por sus siglas en inglés (*extract, transform and load*), busca canalizar los datos que se almacenan en bruto para que luego sean transformados y, finalmente, listos y a disposición. | |
| Informes: son documentos que recogen información de hechos verificados y analizados y que tienen como fin comunicar. | |
| Análisis OLAP: es un concepto y una tecnología para la BI, que permite ejecutar un análisis multidimensional de los datos de una forma rápida e interactiva. Igualmente, ayuda a obtener información resumida de grandes bases de datos u OLTP, con la finalidad de tener análisis previamente calculados, que aceleren la búsqueda de mucha información. El OLAP se emplea principalmente en informes de negocios para las ventas, en *marketing*, en informes de dirección, en minería de datos y áreas similares. | |
| *Data mining*: también conocido en español como *minería de datos*, son aquellos procesos y técnicas a través de los cuales se pueden explorar y comprender las grandes bases de datos y sacarles provecho, con el propósito de encontrar patrones que permitan entender qué pasó, qué está pasando y así predecir lo que pasará en el futuro. | |
| Un sistema de almacén de datos tiene las siguientes características:   * Proporciona la centralización de los activos de datos corporativos. * Está contenido en un entorno bien gestionado. * Tiene definidos procesos consistentes y repetibles para la carga de datos desde aplicaciones corporativas. * Está construido sobre una arquitectura abierta y escalable, que puede manejar la futura expansión de los datos. * Proporciona herramientas que permiten a sus usuarios procesar eficazmente. | |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Carrusel de tarjetas | |
| **Introducción** | Un *data warehouse* se compone de *data marts* y de cubos de información. Ambos se explican de la siguiente manera: | |
| **Fuente:** <https://www.shutterstock.com/es/image-vector/server-room-isometric-cloud-storage-data-179480220>  **Imagen:** 228130\_i12 | | |
| *Data marts:* son almacenadores más pequeños y menos potentes, que contienen subconjuntos de datos. Puede haber uno por cada área o unidad. | | **Figura 5**  *Referente de una data marts*    *Nota.* Reproducidade *Advantages and Disadvantages of a Data Mart. Mbnehaddou. (2020, 16 de enero).* [*https://bit.ly/3gk1WER*](https://bit.ly/3gk1WER)  **Imagen:** 228130\_i13 |
| *Cubos de información:* son bases de datos multidimensionales, que trascienden el esquema bidimensional de las hojas de cálculo. Su interconectividad se facilita gracias al uso de *middleware* o *software* de conexión. | | **Figura 6**  *Referente de cubos de información*    *Nota.* Reproducidade *Tratamiento de los datos: OLTP, OLAP, Data Warehouse. EvaluandoSoftware.com. (2020, 2 de junio).* [*https://bit.ly/3AzJKhh*](https://bit.ly/3AzJKhh)  **Imagen:** 228130\_i14 |

## 1.3. Implementación de bodega de datos

|  |
| --- |
| Cuadro de texto |
| El diseño y la metodología de una bodega de datos debe ser ejecutado de acuerdo con los requerimientos definidos de cada empresa y el modelo de BI que se esté manejando. Además, se deben tener en cuenta los procesos de ETL, que crean el modelo de cómo se aloja la información en el almacén de datos, para que cada usuario tenga acceso a la información correcta. |

# **2. Diseño de bodega de datos**

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Slider Presentación | |
| **Introducción** | El diseño de bodegas de datos comprende una serie de pasos importantes. Es necesario conocer estos pasos, así como la definición del concepto. | |
| Una bodega de datos es una herramienta comercial que se emplea como una solución informática que agrega de diferentes fuentes datos de entidades, archivos u otros sistemas corporativos, y que los convierte, según sus necesidades, alimentándose en un solo archivo. El objetivo de un almacén de datos es facilitar la toma de decisiones corporativas y centralizar, interpretar y agregar valor a los datos, en beneficio de la empresa. Este sistema proporciona a los usuarios la capacidad de analizarlos de forma accesible e intuitiva. Así, un DW es aquel que ha sido extraído y transformado y que está físicamente separado de la fuente de datos. | | **Fuente:** <https://shutr.bz/3Op4MEU>  **Imagen:** 228130\_i15 |
| El desarrollo en el ambiente de la bodega de datos se enfoca en aplicar un proceso de ETL, de tres pasos. El primero es el proceso de *extracción*, que se encarga de obtener datos únicos y necesarios, que, de una u otra manera, están almacenados en diferentes fuentes a la bodega de datos. | | **Fuente:** <https://shutr.bz/3OqIRgH> **Imagen:** 228130\_i16 |
| El siguiente paso es la *transformación*, el cual es el más importante y requiere de mucha precaución, debido a que la confiabilidad de los datos debe ser la mejor. En esta etapa se realiza una limpieza o “modificación” de la información para que se pueda cargar en la bodega. Ambos procesos, limpieza y transformación, se realizan porque los datos, al ser extraídos desde diferentes fuentes, frecuentemente tienen diferentes formatos, lo que los hace poco integrales en la bodega de datos. | | **Imagen:** 228130\_i17 |
| El último paso es la *carga*, que es el almacenamiento del resultado de los pasos anteriormente mencionados; es decir, los datos que fueron extraídos y seguidamente transformados serán cargados en la bodega de datos para un nuevo proceso dentro de la arquitectura de esta bodega.  Los almacenes de datos pueden ser consultados mediante las herramientas de consulta y análisis, como, por ejemplo, cubos multidimensionales, informes o soluciones web. | | **Fuente:** <https://shutr.bz/3Gx3rdB>  **Imagen:** 228130\_i18 |

## 2.1. Sistema operativo y red

|  |  |
| --- | --- |
| **Tipo de recurso** | Cajón de texto de color |
| Antes de empezar, se debe tener en cuenta que estos requerimientos de sistemas operativos y de red no son algo que está establecido ni mucho menos estandarizado, aunque existe una guía básica que puede dar claridad de esto. Del mismo modo, estos requerimientos dependen de la magnitud del proyecto y de la necesidad de los clientes. Para la construcción de una bodega de datos, se deben cumplir algunos requisitos con las herramientas de *software* y *hardware*. En el caso del *hardware*, se requieren uno o más servidores para el almacenamiento de la información y la administración de la base de datos de la organización. Preferiblemente, el servidor debe ser escalable, porque en algunas ocasiones el proyecto de construcción de la bodega presenta un tipo de redimensionamiento a medida que avanza la implementación. | |

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 2 |
| **Introducción** | Inicialmente, la capacidad de almacenamiento se determina por los requerimientos de información histórica de la empresa y su perspectiva de crecimiento. En el caso del *software*, sus herramientas son necesarias para la construcción de bodega de datos, las cuales se clasifican en tres categorías básicas: |
| **Fuente:** <https://shutr.bz/3EOQN8z>  **Imagen:** 228130\_i19 | |
| Herramientas de almacenamiento:se encargan de guardar todos los datos requeridos. Existen diferentes opciones, que dependen de la capacidad del sistema, el volumen de los datos y el presupuesto de la compañía. | |
| Herramientas de extracción: se enfocan en el proceso ETL de los sistemas que usualmente trabaja la compañía. Algunas son:   * SAP BusinessObjects * SQL Server Integration Services * Oracle Data Integrator   Los funcionarios técnicos al interior de la compañía, por su conocimiento en bases de datos y sistemas transaccionales, por lo general, emplean estas herramientas. | |
| Herramientas para la elaboración de reportes*:* son los diferentes programas requeridos para construir la interfaz vista por los usuarios finales. A estos programas se les deben suministrar los recursos necesarios para que puedan visualizar la información requerida de manera detallada y a un nivel gerencial, para que obtengan con ella respuestas a preguntas específicas del negocio. Algunos de estos *softwares* orientados a crear y visualizar reportes son Microsoft Power BI, Looker Studio (antes Google Data Studio), Tableau, entre otros. | |

## 2.2. OLAP

|  |  |
| --- | --- |
| **Tipo de recurso** | Cajón de texto de color |
| El procesamiento analítico en línea (OLAP, por sus siglas en inglés), se refiere a un método para organizar y consultar datos en una estructura multidimensional. Además, a diferencia de los sistemas de datos relacionales, todas las posibles consultas se calculan de antemano, lo que proporciona al usuario de la empresa una mayor agilidad y flexibilidad (Abelló et al., s. f.). | |

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 1 |
| **Introducción** | Los sistemas OLAP constan de ciertas características específicas, que los fabricantes de *software* intentan cumplir en mayor medida y que se basan en las doce leyes que recalcó Edgar F. Codd en 1993. |
| **Fuente:** <https://shutr.bz/3UR5jCj>  **Imagen:** 228130\_i20 | |
| Las características principales de los sistemas OLAP comienzan con integrar una vista conceptual multidimensional, lo que permitirá que se trabaje a partir de métricas de negocio y sus dimensiones. Esto implica al servicio OLAP y a los usuarios aceptar un único esquema lógico de datos. | |
| Por otro lado, la arquitectura cliente/servidor, la dimensionalidad genérica y el rendimiento de informes consistente son características que vale la pena resaltar, debido a que este último no debería degradarse cuando el número de dimensiones del modelo se vaya incrementando. | |
| Asimismo, se debe tener en cuenta que la arquitectura cliente/servidor está basada en sistemas modulares y abiertos, que permitan la interacción y la colaboración. Por su parte, la dimensionalidad genérica se refiere a la capacidad que debe tener el sistema para crear todo tipo de dimensiones y con trabajos que sean proactivos en varias de ellas. | |
| Por último, y no menos importante, se debe aclarar que los sistemas OLAP incumben en el permiso de soporte multiusuario; así que conviene tener una manipulación de datos intuitiva, dado que las personas a las que se destinan estos sistemas son usualmente analistas y altos ejecutivos. Por lo tanto, la interacción debe considerarse desde el prisma de la máxima usabilidad de los usuarios. | |

## 2.3. Característica y entorno técnico del negocio

|  |
| --- |
| Cuadro de texto |
| Los almacenes de datos deben ajustarse a ciertas características, enfocadas en la eficiencia del DW. Según Roldán (2015), la primera está relacionada con utilizar solo los datos necesarios, porque hacer esto será más eficiente que extraer cantidad de información incoherente. Por eso no se debe tomar de las fuentes aquellos datos que no sean requeridos ni menos que no se puedan analizar. La segunda, referida en apartados anteriores, consiste en que la extracción de datos desde diferentes fuentes hace que, en ocasiones, la información no esté integrada. Esta situación produce un pequeño inconveniente, debido a que la integración de los datos ayuda a que la información sea consistente tanto en formato como en atributo. En el caso de que los datos no sean consistentes, se debe realizar el paso dos de un proceso de ETL, es decir, la *transformación* de los datos, con el fin de cumplir oportunamente con la característica. La tercera tiene que ver con la historia, es decir, para realizar posibles análisis de tendencias, se debe tener información en la bodega con datos históricos ―esta es una de las características importantes en este capítulo de diseño de una bodega de datos―. La cuarta característica está relacionada con que no se eliminan ni modifican los datos, sino que se deben realizar dos procesos: la carga de datos y su consulta. En un proyecto real es necesario tener en cuenta criterios de negocio y gestión de proyectos para, por un lado, generar valor en la organización y, por el otro, lograr un proyecto de éxito. |

## 2.4. Requerimientos del usuario

|  |
| --- |
| Cuadro de texto |
| El análisis de los requisitos consiste en identificar las necesidades específicas de una determinada organización en materia de análisis de la información. Los datos recopilados en esta fase permiten señalar posteriormente los procesos empresariales que deben examinarse y las perspectivas organizacionales sobre las que debe evaluarse el rendimiento de la organización.  La lista de requisitos puede ser bastante abrumadora, pero es esencial ponerlos sobre la mesa antes de lanzarse al desarrollo del sistema ETL. Los requerimientos de este sistema son, en su mayoría, limitaciones con las que hay que vivir y a las que hay que adaptarlo. En el marco de estas condiciones, hay oportunidades para tomar sus propias decisiones, ejercer el juicio y aprovechar la creatividad, pero los requisitos dictan los elementos básicos que debe ofrecer el sistema ETL.  Normalmente, en esta fase se debe ser previsor y pensar más allá de las necesidades actuales para satisfacer las futuras. Por ejemplo, si los usuarios han expresado su interés por un análisis mensual y se dispone de información más detallada, es necesario determinar si la solución de factorización de los datos puede beneficiarse de información pormenorizada en el futuro o si son suficientes ciertos niveles de amplificación de la misma. |

# **3. Arquitectura de una bodega de datos**

|  |  |
| --- | --- |
| **Tipo de recurso** | Tarjetas Conectadas |
| **Introducción** | Es importante definir la arquitectura de una bodega de datos. Para ello, se deben analizar las siguientes propiedades esenciales: |
| **Fuente:** <https://shutr.bz/3EPoOFM>  **Imagen:** 228130\_i21 | |
| **Fuente:** <https://shutr.bz/3XhuFL5>  **Imagen:** 228130\_i22 | Separación:los procesos de análisis y transacciones deberían ser almacenados de manera independiente, en la medida de lo posible. |
| **Fuente:** <https://shutr.bz/3OqbUBi>  **Imagen:** 228130\_i23 | Escalabilidad: la infraestructura de *software* y *hardware* debe ser práctica y sencilla de optimizar, de acuerdo con el volumen de los datos. De igual forma, debe estar definida para una fácil administración y gestión de procesos y estimada por el número de solicitudes de los usuarios que determinan aumentos progresivos con el paso del tiempo. |
| **Fuente:** <https://shutr.bz/3XgbTnp>  **Imagen:** 228130\_i24 | Extensibilidad: la arquitectura debe ser capaz de aceptar nuevas aplicaciones y tecnologías, sin tener que rediseñar todo el sistema. |
| **Fuente:** <https://shutr.bz/3XjXkir>  **Imagen:** 228130\_i25 | Seguridad: administrar y verificar los diferentes accesos es una tarea crítica antes de comenzar a realizar el almacenamiento de los datos en un *data warehouse*. |
| **Fuente:** <https://shutr.bz/3ULghsR>  **Imagen:** 228130\_i26 | Administrable:la administración de los datos en la bodega debe ser fácil y sencilla. |

|  |  |
| --- | --- |
| **Tipo de recurso** | Cajón de texto de color |
| Existen dos clasificaciones frecuentemente adoptadas para establecer la arquitectura de una bodega de datos. La primera depende del número de capas que se usa por la arquitectura. La segunda está condicionada por la forma en que cada capa es empleada para crear vistas de la *data warehouse* y estáenfocada en los requerimientos puntuales de una organización o las áreas específicas. | |

## 3.1. Arquitectura de una capa

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 2 |
| **Introducción** | Un primer elemento de consideración en la arquitectura de una bodega de datos es la arquitectura de una capa. En las siguientes líneas se define y caracteriza. |
| **Figura 7**  *Referente arquitectura de una capa*    *Nota.* Reproducidade *Arquitectura de una bodega de datos. FAVA y SENA. (s. f.).* [*https://bit.ly/3tJGrAf*](https://bit.ly/3tJGrAf)  **Imagen:** 228130\_i27 | |
| Una arquitectura de una capa es muy poco usada en la práctica. Su objetivo es reducir la cantidad de los datos guardados eliminando datos redundantes. Esto significa que una bodega de datos se implementa como una vista multidimensional de los datos requeridos y procesados por una capa intermedia específica. La debilidad de este tipo de arquitectura está en su falta de capacidad para satisfacer la necesidad de la independización entre los procesos de análisis y el transaccional. | |
| Las consultas de análisis se envían a los datos operacionales una vez la capa intermedia los interpreta. De esta forma, las consultas realizadas afectan la carga de trabajo de las transacciones normales. | |
| Aunque esta arquitectura puede reunir los requerimientos para integración y exactitud de los datos, no puede agregar más datos que los que vienen de las diferentes fuentes de información iniciales. Por este motivo, un enfoque virtual para el almacenamiento de datos solo puede ser exitoso si los requerimientos de análisis son restringidos y la cantidad de datos a analizar no es muy grande. | |

## 3.2. Arquitectura de dos capas

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tipo de recurso** | Video motion | | | |
| **NOTA** | **La totalidad del texto locutado para el video no debe superar las 500 palabras aproximadamente** | | | |
| **Título** | Arquitectura de dos capas o niveles | | | |
| **Escena** | **Imagen** | **Sonido** | **Narración** | **Texto** |
| **1** | **Fuente:** <https://shutr.bz/3tJnCgP> | Música de fondo | El concepto de *separación* es la esencia en la definición de la arquitectura típica para un sistema de almacenamiento de datos pensado en dos capas. Este tipo de arquitectura corresponde a la separación física entre las fuentes de datos disponibles y lo dispuesto en la bodega, que consta de cuatro etapas posteriores de flujo de información. Estos niveles son los responsables de proporcionar características de recurso, escalabilidad y ganancia, enunciadas a continuación: | Arquitectura de capas o de niveles. |
| **2** | **Fuente:** <https://shutr.bz/3hXVuDX> | Música de fondo | Capa fuente  Un sistema de almacenamiento de datos utiliza fuentes heterogéneas de información. Originalmente, estos se almacenan en la base relacional o bases legadas, o también pueden provenir de sistemas de información externos a la empresa. Esta capa contiene el sistema ETL, el cual se encarga de extraer y validar los datos de las fuentes, convertirlos en información y cargarlos en el suministrador de datos dimensional. | Esta capa está compuesta por diferentes sistemas, que actúan como fuentes de datos. |
| **3** | **Fuente:** <https://shutr.bz/3tJ07UY> | Música de fondo | Extracción de datos  La información existente en las fuentes de datos se extrae, trata y limpia para eliminar cualquier tipo de errores e inconsistencias e integrar fuentes de información heterogéneas. Las herramientas ETL pueden mezclar diferentes fuentes, extraer, transformar, limpiar, validar, filtrar y cargar los datos en una bodega. Esta etapa se encarga de los problemas típicos de los sistemas de información distribuidos, tales como la gestión de datos inconsistentes y estructuras de datos incompatibles. | La extracción de bases de datos es un proceso de recuperación de datos de aquellas bases que son dispares. |
| **4** | **Fuente:** <https://shutr.bz/3TUuZfI> | Música de fondo | *Data warehouse*  Los diferentes datos requeridos se almacenan en un repositorio centralizado, conocido como *data* *warehouse* o bodega de datos. Al DW puede accederse directamente, pero también utilizarse como fuente para la creación de *data marts*, los cuales replican contenidos de los *data warehouse* de manera parcial y se diseñan por requerimientos de áreas específicas en una empresa. Los repositorios que contienen la *meta data* almacenan información, como las fuentes y extracción de datos, usuarios, formas de acceder a ellas, etc. | Es un almacén de datos que se utiliza como repositorio de la información. |
| **5** | **Fuente:** <https://shutr.bz/3ULhdxn> | Música de fondo | Análisis  En la etapa de análisis, la integración de los datos es eficiente y de flexible acceso en la generación de consultas, análisis de información de manera dinámica y simulación de escenarios hipotéticos de negocio. Esta capa funciona con navegadores de datos agregados, optimizadores de consultas complejas e interfaces gráficas de usuario de fácil manejo. Los componentes como *data* *warehouse* son comúnmente denominados DW principal o corporativo y actúan como sistemas de almacenamiento central para todas las consultas que se requieren. | Se analiza la información de forma efectiva para simular proyecciones de acción. |
| **6** | **Fuente:** <https://shutr.bz/3Olqjyk> | Música de fondo | Los siguientes son algunos beneficios de la arquitectura de dos capas:   * La información siempre es utilizable, incluso cuando el acceso a las fuentes no está disponible temporalmente por razones técnicas o alguna política de la organización. * El análisis de la información no afecta el registro de las transacciones. * Se implementan con un modelo multidimensional. * Pueden requerir soluciones especiales de diseño, enfocadas en mejorar la labor de análisis y consultas. | Estos son los beneficios de la arquitectura de dos capas o de niveles. |
| **Nombre del archivo** | 228130\_v2 | | |  |

## 3.3. Arquitectura de tres capas

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 1 |
| **Introducción** | En este tipo de arquitectura, la tercera capa es la de datos reconciliados o almacén de datos operativos. |
| **Figura 8**  *Referente de arquitectura de tres capas*    **Imagen:** 228130\_i28 | |
| La figura anterior muestra una bodega de datos que no es poblada de sus fuentes de forma directa, pero sí a partir de los datos reconciliados.  En esta capa se obtienen los datos operacionales después de realizar las tareas de integración y limpieza de la información de origen. Como resultado, los datos son detallados, coherentes, actuales e integrados. | |
| La ventaja principal de esta capa de reconciliación de datos es que crea un modelo de base de datos común para toda la empresa. | |
| Asimismo, se independizan los problemas de extracción de los datos de las fuentes y la integración de los mismos para llenar la bodega. | |
| A veces esta capa se usa para ejecutar tareas más operativas, como, por ejemplo, la producción de informes diarios, que no pueden ser preparados con las aplicaciones de la empresa. Igualmente crea los flujos adecuados de datos para entregarlos a procesos externos, con el fin de hacer la respectiva limpieza de la información y la integración requerida. | |
| Sin embargo, estos datos conciliados implican una redundancia de información mayor y de origen operativo. | |

## 3.4. *Data marts*

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Tarjetas Avatar | |
| **Introducción** | Estas tarjetas son una agregación de datos contenidos en un almacén primario. Se refiere a subconjuntos de información necesarios para ciertas dependencias del negocio o de roles o tipos de usuarios. | |
| Los *data marts* se generan desde una bodega de datos primaria, y con frecuencia se denominan *dependientes*. Los *data marts* no son estrictamente necesarios, pero sí son útiles para las bodegas de datos de medianas y grandes empresas, por las siguientes razones: | | **Figura 9**  *Referente de arquitectura de tres capas*    *Nota.* Reproducidade *ETL vs. ELT: ¿Cuál es la diferencia? Fátima, N. (2020, 23 de agosto).* [*https://bit.ly/3GuueqH*](https://bit.ly/3GuueqH)  **Imagen:** 228130\_i29 |
| * Se emplean como piezas de construcción a medida que se va incrementando el uso de las bodegas de datos. * Contienen los datos necesarios de determinados grupos de usuarios para generar consultas puntuales. * Ofrecen un rendimiento mejor, ya que son mucho más pequeños que las bodegas de datos primarias. * Algunas veces pueden tener una arquitectura diferente de las fuentes de datos que se usan para llenar los *data marts*. | | **Fuente:** <https://shutr.bz/3Asajor>  **Imagen:** 228130\_i30 |
| En caso de no existir bodega de datos primaria, el proceso de diseño se simplifica, pero podría implicar riesgos de inconsistencias entre los *data marts*. Para evitar estos inconvenientes, podría generarse una bodega de datos principal y crear los *data marts* como independientes. En comparación con la arquitectura de dos capas expuestas, las funciones de estos almacenes de datos y de la bodega estarían invertidas, lo que implica que esta bodega hace recepción mediante los *data marts*, con lo cual puede accederse directamente para generar los patrones de entrada tan sencillos como sea posible. | | **Fuente:** <https://shutr.bz/3gjBv2a>  **Imagen:** 228130\_i31 |
| Las características de una bodega de datos están definidas por   * Empleo de metadatos. Son datos sobre los datos. * Procesos ETL, referidos a la manera en que se almacenan los datos. | | **Fuente:** <https://shutr.bz/3UVdgWT>  **Imagen:** 228130\_i32 |
| Extracción: como su nombre lo indica, corresponde a la obtención de la información desde las diversas fuentes de datos (CRM, ERP, etc.) a una base de datos intermedia, llamada *data staging area* (DSA).  Transformación: es la etapa donde se homogenizan los datos de las diversas fuentes de datos, se limpian y se transforman de acuerdo con los indicadores que se quieran almacenar.  Carga: los datos, una vez depurados y transformados, son cargados en el DW a través de los siguientes procedimientos, que, entre otros, dependen del tipo de aplicación que se otorgue a los datos:   * OLAP * *Query and reporting* * *Executive information system* (EIS) * *Data mining* o minería de datos | | **Figura 10**  *Referente de una arquitectura de bodega de datos*    *Nota.* Reproducidade *Diseño de un data warehouse para medir el desarrollo disciplinar en instituciones académicas. Investigación biotecnológica*, *31(72), 161-181. Gorbea, S. y Madera, M. (2017).* [*https://bit.ly/3EhJ8Og*](https://bit.ly/3EhJ8Og)  **Imagen:** 228130\_i33 |

# **4. Implementación de una bodega de datos**

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Carrusel de tarjetas | |
| **Introducción** | Para llegar a la construcción de una bodega de datos, se necesitan dos tipos de herramientas: *hardware* y *software*. Por un lado, la herramienta *hardware* es un servidor para el almacenamiento y manejo de la base de datos, que debe ser escalable para un redimensionamiento, a medida que se avanza en la implementación. | |
| **Fuente:** <https://bit.ly/3UV8Uz3>  **Imagen:** 228130\_i34 | | |
| La capacidad de almacenamiento debe definirse por los requerimientos y la perspectiva de crecimiento de la organización. En este sentido, es importante:   1. Determinar la misión y los objetivos empresariales. 2. Determinar la funcionalidad que se quiere para el almacén de datos. 3. Determinar los tipos de datos que se necesitan en la bodega para apoyar su funcionalidad. 4. Determinar el contenido en función del volumen y el tamaño. 5. Determinar el tipo de base de datos, puesto que cada una tiene su particularidad. 6. Determinar el diseño físico orientado a producir un buen rendimiento en el procesamiento de consultas. 7. Determinar los procesos de administración futuros del *data warehouse.* 8. Revisar de forma periódica para validar la continua respuesta a las necesidades de la organización. | | **Fuente:** <https://bit.ly/3hXVSlM>  **Imagen:** 228130\_i35 |
| Por otro lado, la herramienta *software* son los programas, aplicaciones o simplemente instrucciones, que permiten realizar un trabajo ofacilitan, optimizan y mejoran la tarea. Para su selección, se deben tener en cuenta diversos aspectos, que están divididos en cuatro categorías básicas: | | **Fuente:** <https://bit.ly/3AwHKqi>  **Imagen:** 228130\_i36 |
| Herramientas de almacenamiento: aquellas en las cuales se almacenan los datos. Dependiendo del volumen de la información, del presupuesto y de la capacidad del sistema, se pueden encontrar varias opciones, como son Oracle, SQL Server, MySQL y PostgreSQL. | | **Fuente:** <https://bit.ly/3i2FGjr>  **Imagen:** 228130\_i37 |
| Herramientas de extracción y colección:realizan el proceso de extracción, transformación y carga de información de los sistemas transaccionales a la bodega de datos. Algunas opciones de herramientas son: BusinessObjects DI, Cognos DM, SQL Server IS, Oracle DI, Data Integration Studio. | | **Fuente:** <https://bit.ly/3tLcHmz>  **Imagen:** 228130\_i38 |
| Herramientas para elaboración de reportes: crean las interfaces vistas por el usuario final. Entre las herramientas más importantes, se encuentran: Microsoft Power BI, Tableau, Google Data Studio, entre otros. | | **Fuente:** <https://bit.ly/3EqLpXD>  **Imagen:** 228130\_i39 |
| Herramientas para construcción de cubos multidimensionales:establecen las medidas y las dimensiones que se usan y genera el cubo para la consulta. Algunas opciones para la construcción de cubos multidimensionales son: Microsoft Analysis Services, Oracle Analytic Workspace Manager y Cognos PowerPlay Transformer. | | **Fuente:** <https://bit.ly/3hZjUgl>  **Imagen:** 228130\_i40 |

## 4.1. Interfases

|  |  |
| --- | --- |
| **Tipo de recurso** | Cajón de texto de color |
| ¿Qué es una interfaz?  Cuando se habla de esta palabra, se pueden encontrar muchos derivativos en el campo de la tecnología, tales como *interfaz de usuario*, *interfaz de programación*, etc. En el ámbito de la bodega de datos, *interfaz* es el conjunto de métodos que deben implementarse de forma obligatoria y que contienen una clase, aunque no tienen por qué comportarse del mismo modo. | |

|  |  |
| --- | --- |
| **Tipo de recurso** | Infografía estática |
| **Texto introductorio** | Es necesario contar con cuatro principios básicos para tener una interfaz exitosa. Estos son los siguientes: |
|  | |
| **Código de la imagen** | 228130\_i41 |

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 2 |
| **Introducción** | Para esta unidad de estudio, se toma como base la metodología Kimball para realizar paso a paso la interfaz de una bodega de datos del área de ventas de una organización. Así, se deben identificar las siguientes tareas a desarrollar: |
| **Fuente:** <https://shutr.bz/3EOhwBW>  **Imagen:** 228130\_i42 | |
| * Establecer el modelo lógico de datos. | |
| * Establecer los procedimientos de conexión entre las fuentes y la bodega de datos. | |
| * Definir el formato de los diferentes datos requeridos. | |
| * Establecer los modelos físicos obtenidos de los modelos lógicos. | |
| * Diseñar los programas y procesos necesarios para los datos y la aplicación. | |
| * Diseñar la bodega de datos. | |
| * Desarrollar los *scripts* que realicen el proceso ETL para poblar la bodega. | |
| * Hacer el montaje, instalación, puesta en marcha y utilización de la bodega de datos. | |
| * Definir las herramientas de acceso del usuario final. | |
| * Diseñar los cubos multidimensionales. | |
| * Generar los cubos multidimensionales. | |
| * Limitar el tiempo de la implementación creando un programa para cumplir con los tiempos, y realizar el respectivo seguimiento. | |
| * Planificar los tiempos de los recursos que serán usados para llevar a cabo las tareas identificadas. | |
| * Hacer seguimiento en la construcción de la bodega mediante un cronograma para identificar posibles incumplimientos en los tiempos previamente establecidos. | |
| * Evitar los tiempos muertos de los recursos y reconocer las tareas adicionales que puedan afectar en el desarrollo e incluirlas dentro del cronograma con sus respectivos tiempos. | |
| * Delegar la responsabilidad de trabajo a los recursos por medio de los tiempos identificados anteriormente. | |
| * Crear un documento final muy específico donde se visualice el plan del proyecto, que incluya cada uno de los puntos mencionados con anterioridad, y donde el equipo encargado de la construcción de la bodega de datos comprenda el mismo enfoque, tenga conocimiento de las actividades y tiempos y evite malas interpretaciones que se puedan evidenciar y que generen retrasos. | |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Slider pasos | |
| **Introducción** | A continuación, se definen los requerimientos generales (cada organización tiene sus propios requerimientos) de este módulo para el área de ventas de una organización: | |
| **Slide 1** | * Procesos que sean medibles. * Información que esté en una sola bodega de datos. * Permitir la integración de distintas fuentes de datos. * Consulta que sea por medio de cubos multidimensionales. | **Figura 11**  *Referente de modelo multidimensional*    *Nota.* Reproducidade *Diseño de una guía general para construir una bodega de datos del área de ventas de una empresa [Tesis de pregrado, Universidad Libre]. Repositorio Unilibre. Roldán, D. E. (2015).* [*https://bit.ly/3UW4ckK*](https://bit.ly/3UW4ckK)  **Imagen:** 228130\_i43 |
| **Slide 2** | Tomando como base el modelo multidimensional propio, se debe diseñar un modelo físico con los nombres de las tablas, campos y tipos de datos: | **Figura 12**  *Referente de un modelo físico*    *Nota.* Reproducidade *Diseño de una guía general para construir una bodega de datos del área de ventas de una empresa [Tesis de pregrado, Universidad Libre]. Repositorio Unilibre. Roldán, D. E. (2015).* [*https://bit.ly/3UW4ckK*](https://bit.ly/3UW4ckK)  **Imagen:** 228130\_i44 |
| **Slide 3** | Se debe contar con un servidor con estas características de *software*: | **Tabla 1**  *Referente de características de un servidor*   |  |  |  |  | | --- | --- | --- | --- | |  | **Requerimientos mínimos para desarrollo y pruebas** | **Requerimientos mínimos para producción** | **Requerimientos recomendados para producción** | | **Arquitectura** | 32bits | 64bits | 64bits | | **Procesador** | 2 núcleos, con velocidad de 2,0 GHz mínimo | 4 núcleos, con velocidad de 2,5 GHz mínimo | 4 núcleos, con velocidad de 2,5 GHz (con Socket libre para un segundo procesador | | **Memoria** | 8GB | 16GB | 32GB | | **Espacio de disco** | 2GB para instalación 5GB para configuraciones y logs | 2GB para instalación 5GB para configuraciones y logs | 2GB para instalación 5GB para configuraciones y logs |   *Nota.* Reproducidade *Diseño de una guía general para construir una bodega de datos del área de ventas de una empresa [Tesis de pregrado, Universidad Libre]. Repositorio Unilibre. Roldán, D. E. (2015).* [*https://bit.ly/3UW4ckK*](https://bit.ly/3UW4ckK)  **Imagen:** 228130\_i45 |
| **Slide 4** | El servidor debe tener las siguientes características de *hardware*: | **Figura 14**  *Referente de un modelo físico*    *Nota.* Reproducidade *Diseño de una guía general para construir una bodega de datos del área de ventas de una empresa [Tesis de pregrado, Universidad Libre]. Repositorio Unilibre. Roldán, D. E. (2015).* [*https://bit.ly/3UW4ckK*](https://bit.ly/3UW4ckK)  **Imagen:** 228130\_i46 |
| **Slide 5** | Los campos para extraerse de cada fuente corresponden a lo siguiente: | **Tabla 2**  *Referente de un modelado dimensional*   |  |  |  |  | | --- | --- | --- | --- | | **Archivos planos cliente** | **Archivos planos productos** | **Tabla empleados** | **Tabla sucursal** | | Id cliente | Id producto | Id empleado | Nombre sucursal | | Nombres cliente | Nombre producto | Nombre empleado | País sucursal | | Apellidos cliente | Descripción producto | Apellidos empleado | Región sucursal | | Sexo cliente | Marca producto | Sexo empleado | Departamento sucursal | | Dirección cliente |  | Dirección empleado | Ciudad sucursal | | Ciudad cliente |  | Ciudad empleado | Dirección sucursal | | Teléfono cliente |  | Teléfono empleado |  | |  |  | Correo empleado |  | |  |  | Cargo |  | |  |  | Sucursal |  |   **Tabla 3**  *Referente de un modelado dimensional*   |  |  |  | | --- | --- | --- | | **Tabla recibos** | **Tabla clientes** | **Tabla productos** | | Fecha | Id cliente | Id producto | | Id recibo | Nombres cliente | Nombre producto | | Id vendedor | Apellidos cliente | Descripción producto | | Id cliente | Sexo cliente | Id marca | | Descripción recibo | Dirección cliente | Nombre marca producto | | Id producto vendido | Ciudad cliente |  | | Cantidad productos | Teléfono cliente |  | | Valor | Profesión cliente |  | |  | Correo cliente |  |   **Imagen:** 228130\_i47 |

## 4.2. Usuarios

|  |  |
| --- | --- |
| **Tipo de recurso** | Rutas / Pasos. Verticales 1 |
| **Introducción** | Cada uno de los usuarios y sus roles son fundamentales para el éxito de una bodega de datos. |
| **Fuente:** <https://shutr.bz/3OlJ98I>  **Imagen:** 228130\_i48 | |
| **Botón 1** | *Project manager*  La responsabilidad de un gestor de proyectos es crear, gestionar y comunicar, tanto verbalmente como por escrito, de manera clara los objetivos al equipo. Debe, además, plantear y defender la ejecución acertada de los pasos para realizar un proyecto y cumplir con los objetivos. También debe tener la habilidad de ser receptivo a las necesidades de los miembros del equipo y dedicarse a la finalización con éxito de todas las tareas del proyecto. |
| **Botón 2** | *Technical leader*  Es la persona que recibe los detalles y los problemas del producto y de la interfaz. Su responsabilidad es asegurarse de que todas las asignaciones encajen y conduzcan a una exitosa implementación del almacenamiento de datos. Trabaja de la mano con el líder del proyecto para lograr el éxito en la implementación. |
| **Botón 3** | *Chief architect*  El arquitecto realiza funciones basadas en la convergencia de las necesidades empresariales, la tecnología informática de almacenamiento de datos actual y las normas y directrices internas de la organización. Su papel en el almacén de datos es creativo, porque tiene que hacer una mirada a la forma en que su empresa puede utilizar la tecnología y los productos para cumplir con los objetivos del negocio o para resolver los problemas que allí surjan. Un arquitecto debe examinar una serie de enfoques y arquitecturas diferentes para luego elegir (o recomendar) la más sensata para ese entorno específico. |
| **Botón 4** | *Data scientist*  Como su nombre lo indica, en una empresa el científico de datos permite que se analicen datos o cualquier otro campo relacionado con el *big data*. En este sentido, facilita la creación de estructuras necesarias en la recolección, análisis y clasificación de los datos para obtener información relevante y, en muchas ocasiones, su visualización. A diferencia del analista, un científico de datos debe saber trabajar con fuentes diferentes y usar la tecnología disponible para el ordenamiento y análisis de las mismas. |
| **Botón 5** | *Data analyst*  El analista de datos es el encargado de unir el departamento de Big Data con el resto de la organización. Si no existe un departamento de esta categoría, entonces esta persona vincula el proyecto de análisis de datos masivos con el resto de la organización. A diferencia del científico de datos, el analista no debe programar o trabajar con múltiples fuentes a la vez, sino que debe estar capacitado para manejar Excel y SQL, incluso puede ir más allá y trabajar con bases más complejas; igualmente, tiene que estar pendiente de los cambios que se puedan producir y ver cuál es la mejor forma de aplicarlos en el campo de trabajo. Este perfil profesional está en la capacidad de ver más allá de los datos: una tendencia, un patrón, una asociación, incluso puede notar datos que aparecen inconexos entre ellos y encontrar una relación en el contexto general. |
| **Botón 6** | *Data strategist*  El rol del estratega de datos puede ser similar al de un analista de datos, con la diferencia de que el primero da respuesta a la fase final del proceso del *big data*. Es la persona que tiene la capacidad analítica y soltura a la hora de entender datos, pero que, sobre todo, tiene mucha creatividad y visión estratégica a partir de la información entregada por el *data* *scientist*. Además, es quien está en contacto directo con el cliente final, lo que le permite percibir las necesidades reales del mercado. |
| **Botón 7** | *Business analyst requirements*  El analista recopila, consolida, organiza y prioriza las necesidades y problemas empresariales que presentan los usuarios. La función de un analista de negocios es utilizar los datos para resolver los problemas empresariales trabajando con el equipo de inteligencia de negocio. |
| **Botón 8** | *Data modeler and conceptual/logical*  Un modelador de datos y diseñador conceptual/lógico es quien organiza una gran cantidad de elementos y crea estructuras de datos que están en sintonía con la forma en que los usuarios accederán a la información y a los tipos de informes y consultas que ejecutarán, tal y como se determinó durante la fase de alcance y las primeras etapas de diseño. |
| **Botón 9** | *Database administrator and physical database designer*  El administrador de bases de datos o diseñador de bases de datos físicas es el encargado degestionar y mantener la base de datos segura y actualizada. Crea sistemas de respaldo para que la información no se pierda si surgen problemas y crea la forma de organizar las bases de datos para hacerlas más rápidas o fáciles de usar. |
| **Botón 10** | *Middleware specialist*  Este especialista se asegura de que los datos se trasladen de forma eficiente y precisa al entorno del almacén de datos, ya sea mediante herramientas o código personalizado. Además, requiere de las funciones de *middleware* de extracción, selección, transformación y otras tareas para convertir los datos de origen en datos listos para el almacén. |

## 4.3. Fuentes de datos

|  |
| --- |
| Cuadro de texto |
| ¿Qué es una fuente de datos?  Es el lugar donde se obtienen datos e informes, que pueden ser una base de datos, un archivo plano, un archivo XML o cualquier otro formato que un sistema pueda leer. Contiene información detallada y que es utilizada en el proceso empresarial para generar análisis que ayudarán a tomar decisiones futuras. Actúa como agente de reestructuración de los datos de la aplicación para hacerla más fácil de consultar e informar.  Sobre la importancia de una fuente de datos, el análisis de esta desempeña un papel fundamental en un proyecto de almacenamiento de información, porque todos los procesos posteriores de extracción y transformación dependen de los datos que contienen estas fuentes, teniendo en cuenta su estructura en cada dato. |

|  |
| --- |
| Cuadro de texto |
| ¡Felicitaciones! En este momento usted ha finalizado el componente formativo con el cual ha aprendido sobre diseño, arquitectura e implementación de las bodegas de bases de datos. Ahora se le invita a revisar la síntesis, donde podrá recordar los saberes que ha desarrollado y realizar la actividad de aprendizaje, donde probará lo aprendido. Podrá, además, consultar el material complementario y el glosario, como refuerzos para profundizar en el conocimiento. |

**Síntesis**

|  |  |
| --- | --- |
| **Tipo de recurso** | Síntesis |
| Tecnólogo en implementación y gestión de bases de datos  Síntesis: Bodegas de datos | |
| **Introducción** | En esta unidad de creación de base de datos, el aprendiz afianzó los conceptos sobre bodega de datos, los cuales necesita para su construcción, acorde con las necesidades de cada organización. Igualmente, esta información le permite reconocer las diferentes opciones de arquitectura que se podrían abordar al momento de definir la creación de estas bases, cómo está organizado el departamento y los principales roles en el equipo con que va a realizar la implementación. Esto con el fin de brindar a la organización de saber diferentes entornos por medio de sus datos históricos, de una manera ágil y con la información necesaria, y así ver el comportamiento interno y tener un mejor control para la toma de decisiones en un futuro.  Como aspecto final de este tema, se le invita a observar el siguiente mapa, que contempla de manera resumida la temática estudiada en el componente formativo: |
| **Figura 15**  *Arquitectura de una bodega de datos*    **Imagen:** 228130\_i49 | |

**Actividad interactiva**

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Actividad didáctica. Opción múltiple | |
| Apreciado aprendiz, a continuación, encontrará diez enunciados relacionados con la experiencia de aprendizaje del presente componente. Seleccione la respuesta correcta de acuerdo con lo aprendido.  ¡Adelante! | | **Fuente:** <https://bit.ly/3AxgDez>  **Imagen 1:** 228130\_i50 |
| 1. Es un sistema de almacenamiento digital que conecta grandes cantidades de datos, de diferentes fuentes: | | **Fuente:** <https://bit.ly/3OmnTQ9>  **imagen 2:** 228130\_i51 |
| a. Bodega de datos (Correcta) | | c. Sistema OLAP |
| b. Storyboard | | d. Dropbox |
| **Retroalimentación**  Respuesta correcta  ¡Muy bien! Ha seleccionado la respuesta correcta.  Respuesta incorrecta  ¡Vamos, inténtelo de nuevo! Lo invitamos a retomar el estudio de los saberes desarrollados en este componente formativo. | | |
| 2. Cuál de las siguientes opciones NO hace parte de los tipos de datos en una bodega de datos: | | **Fuente:** <https://bit.ly/3TV3jaw>  **Imagen 3:** 228130\_i52 |
| a. Datos de funcionamiento de la empresa | | c. Datos de información incoherente y voluminosa (Correcta) |
| b. Integrar los datos del negocio | | d. Datos de seguimiento del negocio |
| **Retroalimentación**  Respuesta correcta  ¡Muy bien! Ha seleccionado la respuesta correcta.  Respuesta incorrecta  ¡Vamos, inténtelo de nuevo! Lo invitamos a retomar el estudio de los saberes desarrollados en este componente formativo | | |
| 3. Cuál de las siguientes opciones NO se relaciona con la finalidad de la bodega de datos: | | **Fuente:** <https://bit.ly/3AvVbq8>  **Imagen 4:** 228130\_i53 |
| a. Almacenar datos de calidad útiles para su uso en los análisis de la empresa | | c. Ayudar al funcionamiento de las aplicaciones que permiten la toma de decisiones |
| b. Mantener un registro completo de las actividades de la empresa | | d. Almacenar todos los datos y registro, sin importar la fuente y su confiabilidad (Correcta) |
| **Retroalimentación**  Respuesta correcta  ¡Muy bien! Ha seleccionado la respuesta correcta.  Respuesta incorrecta  ¡Vamos, inténtelo de nuevo! Lo invitamos a retomar el estudio de los saberes desarrollados en este componente formativo. | | |
| 4. ¿Qué procesos se deben tener en cuenta en el diseño e implementación de una bodega de datos, que creen el modelo de cómo se aloja la información en el almacén de datos? | | **Fuente:** <https://bit.ly/3EYJQ4V>  **Imagen 5:** 228130\_i54 |
| a. Extracción, transformación y carga (Correcta) | | c. Modelado dimensional |
| b. Diagrama de flujo | | d. Modelado tridimensional |
| **Retroalimentación**  Respuesta correcta  ¡Muy bien! Ha seleccionado la respuesta correcta.  Respuesta incorrecta  ¡Vamos, inténtelo de nuevo! Lo invitamos a retomar el estudio de los saberes desarrollados en este componente formativo. | | |
| 5. Es un método para organizar y consultar datos en una estructura multidimensional: | | **Fuente:** <https://bit.ly/3UUSeb3>  **Imagen 6:** 228130\_i55 |
| a. *Data warehouse* | | c. OLAP (Correcta) |
| b. *Data mart* | | d. Ninguna de las anteriores |
| **Retroalimentación**  Respuesta correcta  ¡Muy bien! Ha seleccionado la respuesta correcta.  Respuesta incorrecta  ¡Vamos, inténtelo de nuevo! Lo invitamos a retomar el estudio de los saberes desarrollados en este componente formativo. | | |
| 6. Es un subconjunto o una agregación de los datos guardados en un almacén de datos principal: | | **Fuente:** <https://bit.ly/3gowzJe>  **Imagen 7:** 228130\_i56 |
| a. *Big data* | | c. ETL |
| b. *Data warehouse* | | d. *Data marts* (Correcta) |
| **Retroalimentación**  Respuesta correcta  ¡Muy bien! Ha seleccionado la respuesta correcta.  Respuesta incorrecta  ¡Vamos, inténtelo de nuevo! Lo invitamos a retomar el estudio de los saberes desarrollados en este componente formativo. | | |
| 7.El objetivo de esta arquitectura es minimizar la cantidad de los datos guardados eliminando datos redundantes: | | **Fuente:** <https://bit.ly/3glEtmL>  **Imagen 8:** 228130\_i57 |
| a. Arquitectura mixta | | c. Arquitectura de dos capas |
| b. Arquitectura de tres capas | | d. Arquitectura de una capa (Correcta) |
| **Retroalimentación**  Respuesta correcta  ¡Muy bien! Ha seleccionado la respuesta correcta.  Respuesta incorrecta  ¡Vamos, inténtelo de nuevo! Lo invitamos a retomar el estudio de los saberes desarrollados en este componente formativo. | | |
| 8. El objetivo de esta arquitectura es materializar los datos operativos obtenidos después de la integración y la limpieza de datos desde el origen: | | **Fuente:** <https://bit.ly/3Xwe5HA>  **Imagen 9:** 228130\_i58 |
| a. Arquitectura mixta | | c. Arquitectura de dos capas |
| b. Arquitectura de tres capas (Correcta) | | d. Arquitectura de una capa |
| **Retroalimentación**  Respuesta correcta  ¡Muy bien! Ha seleccionado la respuesta correcta.  Respuesta incorrecta  ¡Vamos, inténtelo de nuevo! Lo invitamos a retomar el estudio de los saberes desarrollados en este componente formativo | | |
| 9. Se encarga de que los datos sean analizables dentro de la empresa y se dedica al análisis de estos o de cualquier otro campo relacionado con el *big data*: | | **Fuente:** <https://bit.ly/3tPSlc6>  **Imagen 10:** 228130\_i59 |
| a. *Data analyst* | | c. *Data scientist* (Correcta) |
| b. *Technical leader* | | d. *Project manager* |
| **Retroalimentación**  Respuesta correcta  ¡Muy bien! Ha seleccionado la respuesta correcta.  Respuesta incorrecta  ¡Vamos, inténtelo de nuevo! Lo invitamos a retomar el estudio de los saberes desarrollados en este componente formativo. | | |
| 10. El objetivo de esta arquitectura es la separación física entre las fuentes de datos disponibles y lo dispuesto en la bodega de datos: | | **Fuente:** <https://bit.ly/3hUOnMt>  **Imagen 11:** 228130\_i60 |
| a. Arquitectura de una capa | | c. Arquitectura de tres capas |
| b. Arquitectura de dos capas (Correcta) | | d. Arquitectura mixta |
| **Retroalimentación**  Respuesta correcta  ¡Muy bien! Ha seleccionado la respuesta correcta.  Respuesta incorrecta  ¡Vamos, inténtelo de nuevo! Lo invitamos a retomar el estudio de los saberes desarrollados en este componente formativo | | |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Actividad didáctica. Verdadero y falso | |
| Apreciado aprendiz, el objetivo principal de la presente actividad es verificar los saberes más importantes del material correspondiente: | | **Fuente:** <https://bit.ly/3tKKIU6>  **Imagen 1:** 228130\_i61 |
| 1. La metodología Inmon y la metodología Kimball son las más usadas en el enfoque de almacenamiento de datos: | | **Fuente:** <https://bit.ly/3USGeGU>  **imagen 2:** 228130\_i62 |
| Verdadero (correcto) | | Falso |
| **Retroalimentación**  Respuesta correcta  ¡Muy bien! Ha seleccionado la respuesta correcta.  Respuesta incorrecta  ¡Vamos, inténtelo de nuevo! Lo invitamos a retomar el estudio de los saberes desarrollados en este componente formativo. | | |
| 2. En una bodega de datos se pueden eliminar o modificar datos para su limpieza y transformación: | | **Fuente:** <https://bit.ly/3ERCkst>  **Imagen 3:** 228130\_i63 |
| Verdadero | | Falso (correcto) |
| **Retroalimentación**  Respuesta correcta  ¡Muy bien! Ha seleccionado la respuesta correcta.  Respuesta incorrecta  ¡Vamos, inténtelo de nuevo! Lo invitamos a retomar el estudio de los saberes desarrollados en este componente formativo. | | |

**Material complementario**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tipo de recurso** | Material complementario | | |
| Tema | Referencia APA del material | Tipo | Enlace |
| *Data warehouse* | Conesa, J. y Curto, J. (2015). ¿*Cómo crear un data warehouse?* Editorial UOC. | Libro | <https://www.editorialuoc.cat/como-crear-un-data-warehouse> |
| Implementación de una bodega de datos | Conexión ESAN*.* (2020, 1 de julio). Data warehouse: pasos para construirlo con éxito. *ESAN Business*. | Página web | <https://www.esan.edu.pe/conexion-esan/data-warehouse-pasos-para-construirlo-con-exito> |
| Ingeniería de *software* | Echeverri, J., Aristizábal, M., González, L., Urrego, G., Polo, R., et al. (2013). *Reflexiones sobre ingeniería de requisitos y pruebas de software*. (2013). Corporación Universitaria Remington; Organización LACREST. | Libro | <https://www.uniremington.edu.co/wp-content/uploads/2019/01/LACREST-2013-uniremington.pdf> |
| Ciencia de datos | García, J., Berlanga, A., Molina, J. M., Patricio M. A., Bustamante A. L. y Padilla, W. R., (2018). *Ciencia de datos. Técnicas analíticas y aprendizaje estadístico*. Altaria. | Libro | <https://dialnet.unirioja.es/servlet/libro?codigo=763464> |
| Estructura secuencial | GSL Industrias. (2021, 7 de junio). *Estructura secuencial.* | Página web | <https://industriasgsl.com/blogs/automatizacion/que-es-la-estructura-secuencial> |
| Programación | Lewis, R. M. (2017). *Tutorial fácil de seguir para aprender la programación de Python en menos de una semana*. | Libro digital | <https://bit.ly/3glAJ4w> |
| *Data warehouse* | Luna, E. y García, F. (2000). El repositorio de metadatos en un *data warehouse*. *Revista Facultad de Ingeniería*, (8), 10-15. | Revista | <https://www.redalyc.org/pdf/114/11400802.pdf> |
| Bases de datos relacionales | Martínez, F. J. y Gallegos, A. (2017). *Programación de bases de datos relacionales*. DELAU. | Libro | <https://librerianacional.com/producto/350241> |
| *Software* | Medina, J., Pineda, E. y Téllez, F. R. (2019). Requerimientos de *software*: prototipado, *software* heredado y análisis de documentos. *Ingeniería y Desarrollo*, *37*(2), 327-345. | Revista | <https://rcientificas.uninorte.edu.co/index.php/ingenieria/article/view/11452/214421444623> |
| Elementos de una bodega de datos | SAP. (s. f.). *¿Qué es un almacén de datos?* | Página web | <https://www.sap.com/latinamerica/insights/what-is-a-data-warehouse.html> |
| Sistema de bases de datos | Silberschatz, A., Korth, H. F. y Sudarshan, S. (2019). *Database System Concepts* (7.a ed.). McGraw-Hill. | Libro | <https://www.db-book.com/> |
| Bases de datos | Wanumen, L. F., Rivas, E. y Mosquera, D. J. (2018). *Bases de datos en SQL server*. Ecoe Ediciones. | Libro | <https://www.buscalibre.com.co/libro-bases-de-datos-en-sql-server/9789587715705/p/50093560> |

**Glosario**

|  |  |
| --- | --- |
| **Tipo de recurso** | Glosario |
| *Business intelligence* (BI): | método que agrupa los datos de una empresa, con el objetivo que se analicen y transformen en información que pueda utilizarse para mejorar los procesos. |
| *Data mart*: | o mercado de datos en español, es un subconjunto de datos orientado a un objeto. Su principal característica es su diseño clásico, que facilita en gran medida su uso (PowerData, 2016). |
| *Data mining*: | o minería de datos, es el proceso técnico que analiza grandes cantidades de información dispersa para darle sentido y convertirla en conocimiento (KeepCoding, 2022a). Es la forma de descubrir información relevante. |
| ETL: | extracción, transformación y carga de datos. |
| Metadatos: | son datos sobre los datos, que explican el tipo de información que se contiene en cada campo de las tablas. |
| MOLAP: | *Multi-dimensional online* *analytical processing*, en español: procesamiento analítico multidimensional en línea, es una herramienta OLAP cuyo sistema guarda los datos en una matriz multidimensional de almacenamiento y requiere que el procesamiento y la acumulación de información estén contenidos en el cubo OLAP (Evaluando Software, 2022). |
| ODS: | *Operational data store*, o almacén operacional de datos, es un contenedor de datos activos, diseñado para integrar información de múltiples fuentes (KeepCoding, 2022b). |
| OLAP: | el procesamiento analítico en línea es un conjunto de herramientas de *software*, que proporciona un marco tridimensional para la toma de decisiones. |
| ROLAP: | *Relational online analytical processing*, en español: procesamiento analítico en línea relacional, es una herramienta OLAP construida sobre una base de datos relacional. En este sistema tiene importancia la tabla de hechos, donde se almacena la historia de la información relevante para la empresa que requiere ser estudiada (Evaluando Software, 2022). |
| SQL: | El lenguaje de consulta estructurado permite a los usuarios definir la estructura y organización de los datos almacenados y, a su vez, verifica y mantiene su integridad y control de acceso, con lo cual deja definidas las relaciones entre los elementos de datos almacenados (PowerData, 2016). |

**Referencias bibliográficas**

|  |  |
| --- | --- |
| **Tipo de recurso** | Bibliografía |
| Abelló, A., Curto, J., Rius, À., Serra, M., Samos, J. y Vidal, J. (s. f.). *Introducción al Data Warehouse*. UOC. <https://bit.ly/3Ay8Woh> | |
| Bahamón, J. H. (2003). Propuesta de un método para el diseño y modelado de una bodega de datos. *Sistemas & Telemática*,13-37. <https://www.icesi.edu.co/contenido/pdfs/sistemas_telematica2.pdf> | |
| Biscobing, J. (2021). *Almacén de datos (data warehouse).* <https://www.computerweekly.com/es/definicion/Almacen-de-datos-data-warehouse> | |
| Evaluando Software. (2022, 21 de febrero). *ABC del OLAP.* <https://www.evaluandosoftware.com/abc-del-olap/> | |
| KeepCoding. (2022a, 27 de enero). *¿Qué es el data mining o minado de datos?* <https://keepcoding.io/blog/que-es-el-data-mining/> | |
| KeepCoding. (2022b, 1 de junio). *Glosario de Data Warehouse: 7 conceptos fundamentales*. <https://keepcoding.io/blog/glosario-data-warehouse-conceptos-fundamentales/> | |
| Naeem, T. (2020, 3 de febrero). Conceptos de *Data Warehouse*: enfoque de Kimball vs. Inmon. *Astera*. <https://www.astera.com/es/type/blog/data-warehouse-concepts/> | |
| PowerData. (2016, 9 de abril). *Data warehouse concepts: tu glosario de almacén de datos*. <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/data-warehouse-concepts-tu-glosario-de-almacen-de-datos> | |
| Roldán, D. (2015). *Diseño de una guía general para construir una bodega de datos del área de ventas de una empresa* [Tesis de pregrado, Universidad Libre]. Repositorio Institucional Unilibre. <https://bit.ly/3UYtKxy> | |
| SAS. (s. f.). *Data Warehouse. Qué es y por qué es importante*. <https://www.sas.com/es_pe/insights/data-management/data-warehouse.html> | |