



TECNOLÓGICO
NACIONAL DE MÉXICO



TECNOLÓGICO NACIONAL DE MÉXICO
INSTITUTO TECNOLÓGICO DE CIUDAD MADERO



TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE CIUDAD MADERO

Carrera: Sistemas Computacionales

Tema: Arquitectura de base de datos

Alumno: Reyes Villar Luis Ricardo

Numero de control: 21070343

Profesora: Elizabeth Cortez Razo

Materia: Fundamentos de Base de datos

Hora: 9:00 – 10:00hrs

Grupo: 4504A

Fecha de entrega: 07 de Febrero del 2023

Semestre: Enero 2023 – Junio 2023

¿Qué es la arquitectura de base de datos?

Una arquitectura de base de datos o arquitectura de datos describe cómo se gestionan los datos, desde la recopilación hasta la transformación, la distribución y el consumo. Establece el plan para los datos y la forma en que fluyen a través de los sistemas de almacenamiento de datos. Es fundamental para las operaciones de procesamiento de datos y las aplicaciones de inteligencia artificial (IA).

Una arquitectura de datos define los procesos de captura, transformación y entrega de datos utilizables a los usuarios de una empresa. Y lo que es más importante, identifica a las personas que van a consumir esos datos y sus requisitos únicos. Una buena arquitectura de datos fluye de derecha a izquierda: de los consumidores de datos a las fuentes de datos.

El diseño de una arquitectura de datos debe regirse por los requisitos del negocio, que los arquitectos de datos y los ingenieros de datos utilizan para definir su respectivo modelo de datos y las estructuras de datos subyacentes que lo sustentan. Estos diseños suelen responder a una necesidad de negocio.

Una arquitectura de datos debe establecer estándares de datos para todos sus sistemas de datos como una visión o un modelo de las eventuales interacciones entre esos sistemas de datos.

Una arquitectura de datos describe las estructuras de datos utilizadas por una empresa y su software de aplicaciones informáticas. La arquitectura de datos aborda los datos almacenados, los datos en uso y los datos en movimiento; las descripciones de los almacenes de datos, los grupos de datos y los elementos de datos; y las asignaciones de esos artefactos de datos a las calidades de datos, las aplicaciones, las ubicaciones, etc. Proporciona criterios para las operaciones de procesamiento de datos que permiten diseñar flujos de datos y también controlar el flujo de datos en el sistema.

Es el proceso de estandarización de cómo las organizaciones recogen, almacenan, transforman, distribuyen y utilizan los datos.

¿Cuál es el objetivo de la arquitectura de base de datos?

Una arquitectura de datos orienta a los organismos del Gobierno a realizar una eficiente gestión de datos y propone un esquema de componentes compuestos por herramientas, buenas prácticas y estándares. Es decir, el objetivo es ofrecer datos relevantes a las personas que los necesitan, cuando los necesitan, y ayudarles a darles sentido.

Descripción de la arquitectura de base de datos.

Una buena arquitectura de datos garantiza que los datos sean manejables y útiles, lo que facilita la gestión del ciclo de vida.

Una arquitectura de datos identifica los procesos, sistemas y operativa necesaria para identificar, acceder, interoperar y gestionar los datos de gobierno.

Dentro de una arquitectura de datos se desarrollan los siguientes elementos:

- Modelo de Datos: Entidades y Vocabularios
- Principios que rigen a los datos de gobierno
- Mapa conceptual
- Ciclo de vida de los datos
- Análisis de la cadena de valor de información y entrega de datos
- Diagrama lógico de datos, etc

Una arquitectura de datos también comprende:

- Entender las necesidades de información, para poder desarrollar y mantener el modelo de datos;
- Definir y mantener la arquitectura de Base de Datos;
- Definir y mantener la arquitectura de Integración de Datos;
- Definir y mantener la arquitectura de Big Data y Analítica;
- Definir y mantener la arquitectura de Metadatos.

Componentes de arquitectura de datos.

La arquitectura de datos puede sintetizarse en tres componentes generales:

- **Resultados de arquitectura de datos.** Son los modelos, las definiciones y los flujos de datos que suelen denominarse artefactos de arquitectura de datos.
- **Actividades de arquitectura de datos.** Son las formas, despliegues y cumplimientos de las intenciones de la arquitectura de datos.
- **Comportamientos de arquitectura de datos.** Son las colaboraciones, mentalidades y habilidades de los distintos roles que afectan a la arquitectura de datos de una empresa.

Características de una arquitectura de base de datos.

La arquitectura de datos se construye en torno a ciertas características:

- ❖ **Automatización.** Los procesos que tardaban meses en construirse ahora pueden completarse en horas o días utilizando herramientas basadas en la nube. Si un usuario quiere acceder a diferentes datos, la automatización permite al arquitecto diseñar rápidamente una canalización para entregarlos. A medida que se obtienen nuevos datos, los arquitectos de datos pueden integrarlos rápidamente en la arquitectura. Y para crear una arquitectura adaptable en la que los datos fluyan continuamente, los arquitectos de datos lo automatizan todo.
- ❖ **Seguridad.** La seguridad está integrada en la arquitectura de datos moderna, lo que garantiza que los datos estén disponibles en función de la necesidad de conocimiento definida por la empresa. Una buena arquitectura de datos también reconoce las amenazas existentes y emergentes para la seguridad de los datos, y garantiza el cumplimiento normativo de la legislación como HIPAA y GDPR.
- ❖ **Orientación al usuario.** En la arquitectura de datos moderna, los usuarios de la empresa pueden definir con confianza los requisitos, ya que los arquitectos de datos pueden agrupar los datos y crear soluciones para

acceder a ellos de forma que cumplan los objetivos empresariales. Una buena arquitectura de datos evoluciona continuamente para satisfacer las nuevas y cambiantes necesidades de información de los clientes.

- ❖ **Resiliente.** Cualquier arquitectura de datos debe ser resiliente, con alta disponibilidad, recuperación de desastres y capacidades de copia de seguridad/restauración.
- ❖ **Canalizaciones de datos escalables.** Para aprovechar las tecnologías emergentes, las arquitecturas de datos admiten el flujo de datos en tiempo real y las ráfagas de datos en microlotes.
- ❖ **Colaboración.** Una buena arquitectura de datos elimina los silos combinando los datos de todas las partes de la organización, junto con las fuentes externas que sean necesarias, en un solo lugar para eliminar las versiones competitivas de los mismos datos. En este entorno, los datos no se intercambian entre las unidades de negocio ni se acumulan, sino que se consideran un activo compartido por toda la empresa.
- ❖ **Impulsada por la IA.** La arquitectura de datos utiliza el aprendizaje automático y la inteligencia artificial para construir los objetos de datos, las tablas, las vistas y los modelos que mantienen el flujo de datos. La arquitectura de datos inteligente lleva la automatización a un nuevo nivel, utilizando el aprendizaje automático (ML) y la inteligencia artificial (AI) para ajustar, alertar y recomendar soluciones a las nuevas condiciones. El ML y la IA pueden identificar los tipos de datos, identificar y corregir los errores de calidad de los datos, crear estructuras para los datos entrantes, identificar las relaciones para obtener nuevos conocimientos y recomendar conjuntos de datos y análisis relacionados.
- ❖ **Elasticidad.** La elasticidad permite a las empresas ampliar o reducir la escala según sea necesario. La nube permite la escalabilidad bajo demanda de forma rápida y asequible. La elasticidad permite a los administradores centrarse en la resolución de problemas. Las arquitecturas elásticas liberan a los administradores de tener que calibrar la capacidad con exactitud, estrangular el uso si es necesario y comprar hardware en exceso sin cesar.

La elasticidad también da lugar a muchos tipos de aplicaciones y casos de uso, como entornos de desarrollo y prueba bajo demanda, cajas de arena analíticas y zonas de juego para la creación de prototipos.

- ❖ **Simplicidad.** La simplicidad supera a la complejidad en una arquitectura de datos eficiente. La simplicidad en el movimiento de datos, las plataformas de datos, los marcos de montaje de datos y las plataformas analíticas son importantes aportes en una arquitectura de datos. Para reducir la complejidad, las organizaciones deben esforzarse por limitar el movimiento y la duplicación de datos y abogar por una plataforma de base de datos, un marco de ensamblaje de datos y una plataforma analítica uniformes
- ❖ **Adaptable.** Una arquitectura de datos moderna debe ser lo suficientemente flexible como para soportar una multiplicidad de necesidades empresariales. Tiene que soportar múltiples tipos de usuarios empresariales, operaciones de carga y tasas de actualización, operaciones de consulta, despliegues, motores de procesamiento de datos y pipelines.
- ❖ **Gobernanza.** Una arquitectura de datos moderna define puntos de acceso para que cada tipo de usuario pueda satisfacer sus necesidades de información. Los científicos de datos deben tener acceso a los datos brutos en el área de aterrizaje o, mejor aún, a un sandbox construido a propósito donde puedan mezclar datos corporativos brutos con sus propios datos.
- ❖ **Nube nativa.** Las arquitecturas de datos modernas están diseñadas para soportar el escalado elástico, la alta disponibilidad, la seguridad de extremo a extremo para los datos en movimiento y los datos en reposo, y la escalabilidad de costes y rendimiento.
- ❖ **Integración de datos sin fisuras.** Las arquitecturas de datos se integran con las aplicaciones heredadas mediante interfaces API estándar. Están optimizadas para compartir datos entre sistemas, geografías y organizaciones.
- ❖ **Habilitación de datos en tiempo real.** Las arquitecturas de datos modernas admiten la capacidad de desplegar la validación, la clasificación, la gestión y el gobierno de los datos de forma automatizada y activa.

- ❖ **Desacopladas y extensibles.** Las arquitecturas de datos modernas están diseñadas para que estén poco acopladas, permitiendo que los servicios realicen tareas mínimas independientes de otros servicios.

Comparación entre modelos de datos conceptuales, lógicos y físicos.

La documentación de la arquitectura de datos incluye tres tipos de modelos de datos.

- **Modelos de datos conceptuales:** también conocidos como modelos de dominio, ofrecen una imagen general de lo que contendrá el sistema, cómo se organizará y qué reglas de negocio se aplican. Los modelos conceptuales normalmente se crean como parte del proceso de recopilación de los requisitos iniciales del proyecto. Por lo general, incluyen clases de entidad (que definen los tipos de cosas que son importantes para representar el negocio en el modelo de datos), sus características y restricciones, las relaciones entre ellas y los requisitos relevantes de seguridad e integridad de los datos.
- **Modelos de datos lógicos:** son menos abstractos y brindan más detalles sobre los conceptos y las relaciones en el dominio bajo consideración. Se sigue uno de los sistemas de notación de modelado de datos formales. Estos indican atributos de datos, como tipos de datos y sus longitudes correspondientes, y muestran las relaciones entre entidades. Los modelos de datos lógicos no especifican ningún requisito técnico del sistema.
- **Modelos de datos físicos:** son menos abstractos y brindan más detalles sobre los conceptos y las relaciones en el dominio bajo consideración. Se sigue uno de los sistemas de notación de modelado de datos formales. Estos indican atributos de datos, como tipos de datos y sus longitudes correspondientes, y muestran las relaciones entre entidades. Los modelos de datos lógicos no especifican ningún requisito técnico del sistema.

Marcos de trabajo populares de arquitectura de base de datos.

Una arquitectura de datos puede basarse en marcos de trabajo populares de arquitectura empresarial, incluyendo TOGAF, DAMA-DMBOK 2 y Zachman Framework for Enterprise Architecture.

The Open Group Architecture Framework (TOGAF)

Esta metodología de arquitectura empresarial fue desarrollada en 1995 por The Open Group.

Esta arquitectura se basa en cuatro pilares:

- **Arquitectura de negocio**, que define la estructura organizativa, la estrategia de negocio y los procesos de la empresa.
- **Arquitectura de datos**, que describe los activos de datos físicos, lógicos y conceptuales y cómo se almacenan y gestionan a lo largo del ciclo de vida.
- **Arquitectura de aplicaciones**, que representa los sistemas de aplicaciones, y cómo se relacionan con los principales procesos de negocio y entre ellos.
- **Arquitectura técnica**, que describe la infraestructura tecnológica (hardware, software y redes) necesaria para dar soporte a las aplicaciones más importantes.

Como tal, TOGAF proporciona un marco de trabajo completo para diseñar e implementar la arquitectura de TI de una empresa, incluida su arquitectura de datos.

DAMA-DMBOK 2

DAMA International, originalmente fundada como Data Management Association International, es una organización sin ánimo de lucro dedicada al avance en la gestión de datos y la información. Su cuerpo de conocimiento de gestión de datos, DAMA-DMBOK 2, cubre la arquitectura de datos, así como el gobierno y la ética, el modelado y el diseño de los datos, el almacenamiento, la seguridad y la integración.

Zachman Framework for Enterprise Architecture

Desarrollado originalmente por John Zachman en IBM en 1987, este marco de trabajo utiliza una matriz de seis capas, desde la contextual hasta la detallada, correlacionadas con seis preguntas que incluyen por qué, cómo y qué. Proporciona un modo formal de organizar y analizar datos, pero no incluye métodos para hacerlo.

Tipos de arquitecturas base de datos y componentes subyacentes.

Una arquitectura de datos demuestra una perspectiva de alto nivel de cómo trabajan juntos los diferentes sistemas de gestión de datos. Estos sistemas incluyen diferentes repositorios de almacenamiento de datos, tales como lagos de datos, almacenes de datos, data marts o bases de datos. Juntos, pueden crear arquitecturas de datos, como entramados de datos y mallas de datos. Estas arquitecturas se centran más en los datos como productos, favoreciendo la estandarización en torno a metadatos y la democratización de los datos en las distintas organizaciones a través de las API.

En la siguiente sección se describe más detenidamente cada uno de estos componentes de almacenamiento y tipos de arquitectura de datos:

Tipos de sistemas de gestión de base de datos

Almacenes de datos: un depósito de datos agrega datos de los diferentes orígenes de toda una empresa en un repositorio único, central y coherente. Después de la extracción, los datos fluyen a través de una canalización de datos ETL, pasando por varias transformaciones para cumplir con el modelo de datos predefinido. Una vez que se cargan en el depósito de datos, los datos pasan a dar soporte a diferentes aplicaciones de inteligencia empresarial (BI) y ciencia de datos.

Data marts: un data mart es una versión focalizada de un depósito de datos que contiene un subconjunto más pequeño de los datos importantes y necesarios para un equipo individual o un grupo seleccionado de agrupar dentro de una organización, como el departamento de recursos humanos. Dado que contienen un subconjunto de datos más pequeño, los data marts permiten que un departamento

o una línea de negocio descubran información más focalizada de forma más rápida que cuando se trabaja con el conjunto más amplio de datos del depósito de datos.

Lagos de datos: mientras que los almacenes de datos procesan datos, un lago de datos alberga datos en bruto, normalmente petabytes. Un lago de datos puede almacenar tanto datos estructurados como no estructurados, lo que lo diferencia de otros repositorios de datos. Esta flexibilidad en los requisitos de almacenamiento resulta especialmente útil para los científicos de datos, los ingenieros de datos y los desarrolladores, ya que les permite acceder a los datos para ejercicios de descubrimiento de datos y proyectos de machine learning. Si bien los lagos de datos son más lentos que los almacenes de datos, también son más baratos, ya que apenas requieren preparación de datos antes de la ingesta. Actualmente, siguen evolucionando como parte de los esfuerzos de migración de datos al cloud. Los lagos de datos admiten una amplia variedad de casos de uso, ya que no es necesario definir los objetivos de negocio para los datos en el momento de su recopilación. Sin embargo, dos casos de uso principales sí incluyen la exploración de la ciencia de datos y las tareas de copia de seguridad y recuperación de datos. Los científicos de datos pueden utilizar lagos de datos para pruebas de concepto. Las aplicaciones de machine learning se benefician de la capacidad de almacenar datos estructurados y no estructurados en el mismo sitio, lo que no es posible utilizando un sistema base de datos relacional. Los lagos de datos también se pueden utilizar para probar y desarrollar proyectos de análisis de big data. Cuando se ha desarrollado la aplicación y se han identificado los datos útiles, los datos se pueden exportar a un depósito de datos para su uso operativo, y se puede aplicar automatización para escalar la aplicación. Los lagos de datos también se pueden utilizar para la copia de seguridad y recuperación de datos, debido a su escalabilidad a bajo coste. Por las mismas razones, los lagos de datos son buenos para almacenar datos "por si acaso", para los cuales aún no se han definido necesidades del negocio. Almacenar los datos ahora significa que estarán disponibles más adelante cuando vayan surgiendo nuevas iniciativas.

Tipos de arquitecturas de datos

Entramados de datos: un entramado de datos es una arquitectura, que se centra en la automatización de la integración de los datos, la ingeniería de datos y el gobierno en una cadena de valor de los datos entre sus proveedores y consumidores. Un entramado de datos se basa en la noción de "metadatos de activos" que utiliza gráficos de conocimientos, semántica, minería de datos y tecnología de machine learning (ML) para descubrir patrones en varios tipos de metadatos (por ejemplo, registros del sistema, redes sociales, etc.). A continuación, aplica toda esta información de valor para automatizar y orquestar la cadena de valor de datos. Por ejemplo, puede permitir que un consumidor de datos encuentre un producto de datos y suministrarle ese producto de datos automáticamente. Un mayor acceso a los datos entre los productos de datos y sus consumidores permite reducir los silos de datos y proporciona una imagen más completa de los datos de la organización. Los entramados de datos son una tecnología emergente con un enorme potencial y se pueden utilizar para mejorar la creación de perfiles de clientes, la detección de fraude y el mantenimiento preventivo. Según Gartner, los entramados de datos reducen el tiempo de diseño de la integración en un 30 %, el tiempo de despliegue en un 30 % y el mantenimiento en un 70 %.

Mallas de datos: una malla de datos es una arquitectura de datos descentralizada que organiza los datos por dominio empresarial. Al usar una malla de datos, la organización tiene que dejar de pensar en los datos como un producto secundario de un proceso y empezar a considerarlos como un producto en sí mismo. Los productores de datos actúan como propietarios del producto de datos. Como expertos en la materia, los productores de datos pueden utilizar sus conocimientos de los consumidores de datos primarios para diseñar API para ellos. Desde otras partes de la organización también se podrá acceder a estas API, ampliando así el acceso a los datos gestionados.

Los sistemas de almacenamiento más tradicionales, como los lagos de datos y los almacenes de datos, se pueden utilizar como varios repositorios de datos descentralizados para realizar una malla de datos. Una malla de datos también

puede funcionar con un entramado de datos, con la automatización del entramado de datos que permite crear nuevos productos de datos más rápidamente o imponer el gobierno global.

Ventajas de las arquitecturas de base de datos.

Una arquitectura de datos bien construida puede ofrecer a las empresas distintas ventajas clave, que incluyen:

Reducción de la redundancia: puede haber campos de datos que se solapen en diferentes orígenes, lo que incrementa el riesgo de incoherencia, imprecisiones de datos y oportunidades perdidas para la integración de los datos. Una buena arquitectura de datos puede estandarizar el modo en que se almacenan los datos y potencialmente reducir la duplicación, lo que favorece una mejor calidad y análisis integrales.

Mejora de la calidad de los datos: las arquitecturas de datos bien diseñadas pueden resolver algunos de los desafíos de los lagos de datos mal gestionados, también conocidos como "pantanos de datos". Un pantano de datos carece de prácticas apropiadas de calidad y gobierno de datos para proporcionar conocimientos prácticos. Las arquitecturas de datos pueden ayudar a imponer el gobierno de datos y estándares de seguridad de datos, con lo cual se puede llevar a cabo la supervisión adecuada en la canalización de datos para garantizar su funcionamiento según lo previsto. Al mejorar la calidad y el gobierno de los datos, las arquitecturas de datos aseguran que los datos se almacenen de manera que resulten útiles ahora y en el futuro.

Habilitación de la integración: los datos a menudo están aislados, a causa de limitaciones técnicas en el almacenamiento de datos y barreras organizativas dentro de la empresa. Las arquitecturas de datos actuales deben centrarse en facilitar la integración de los datos entre dominios, de modo que diferentes zonas geográficas y funciones de negocio tengan acceso a los datos de los demás. Esto permite comprender mejor las métricas comunes (como gastos, ingresos y sus factores

asociados), además de ofrecer una visión más integral de los clientes, los productos y las zonas geográficas, para tomar mejores decisiones.

Gestión del ciclo de vida de los datos: una arquitectura de datos moderna puede dirigir el modo de gestionar los datos a lo largo del tiempo. Los datos suelen perder utilidad a medida que envejecen y se accede a ellos con menos frecuencia. Con el tiempo, los datos se pueden migrar a tipos de almacenamiento más baratos y lentos, de manera que sigan estando disponibles para informes y auditorías, pero sin el gasto de un almacenamiento de alto rendimiento.

Arquitectura de base de datos moderna.

A medida que las organizaciones establecen su hoja de ruta para las aplicaciones futuras, incluidas las cargas de trabajo de IA, blockchain e Internet de las cosas (IoT), necesitan una arquitectura de datos moderna capaz de atender los requisitos de los datos.

Las siete características principales de una arquitectura de datos moderna son:

1. Nativa en cloud y habilitada para cloud, para que la arquitectura de datos pueda beneficiarse del escalado elástico y la alta disponibilidad del cloud.
2. Canalizaciones de datos sólidas, escalables y portátiles, que combinen flujos de trabajo inteligentes, analítica cognitiva e integración en tiempo real en un único marco de trabajo.
3. Integración de datos perfecta, utilizando interfaces de API estándar para conectarse a las aplicaciones existentes.
4. Habilitación de datos en tiempo real, que incluye la validación, la clasificación, la gestión y el gobierno.
5. Desacoplada y extensible, por lo que no existen dependencias entre servicios y los estándares abiertos facilitan la interoperatividad.
6. Basada en dominios de datos comunes, sucesos y microservicios.
7. Optimizada para equilibrar coste y simplicidad.

Principios de la arquitectura de base de datos.

Según Joshua Klahr, vicepresidente de AtScale, hay seis principios que forman la base de la arquitectura de datos moderna:

1. **Los datos son un activo compartido.** Una arquitectura de datos moderna debe eliminar los silos de datos departamentales y ofrecer a todas las partes interesadas una visión completa de la empresa.
2. **Los usuarios necesitan un acceso adecuado a los datos.** Las arquitecturas de datos modernas deben ofrecer interfaces que faciliten a los usuarios el consumo de datos mediante herramientas adecuadas a su trabajo.
3. **La seguridad es esencial.** Las arquitecturas de datos modernas deben estar diseñadas para la seguridad y deben soportar políticas de datos y controles de acceso directamente en los datos brutos.
4. **Los vocabularios comunes garantizan una comprensión común.** Los activos de datos compartidos, como los catálogos de productos, las dimensiones del calendario fiscal y las dimensiones de los KPI, requieren un vocabulario común para ayudar a evitar disputas durante el análisis.
5. **Los datos deben ser rehabilitados.** Invertir en funciones básicas que realicen el cuidado de datos.
6. **Los flujos de datos deben ser optimizados para la agilidad.** Reduzca el número de veces que los datos deben ser movidos para reducir el coste, aumentar la frescura de los datos y optimizar la agilidad de la empresa.

Referencias Bibliográficas.

Que es la arquitectura de datos. (2020). gub.uy. Recuperado de <https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/politicas-y-gestion/arquitectura-datos#:~:text=La%20Arquitectura%20de%20Datos%20identifica,a%20los%20datos%20de%20gobierno>

¿Qué es una arquitectura de datos? (2022). IBM. Recuperado de <https://www.ibm.com/es-es/topics/data-architecture>

Novoseltseva Ekaterina. Características y principios de las bases de datos. (2021). Apiumhub. Recuperado de <https://apiumhub.com/es/tech-blog-barcelona/caracteristicas-principios-arquitectura-de-datos/>