

# Data Management Fundamentals

## Módulo 14 – Gestión de Metadatos

Fernando Gualo Cejudo

### GOLD SPONSORS



### SILVER SPONSORS



### Entidades académicas



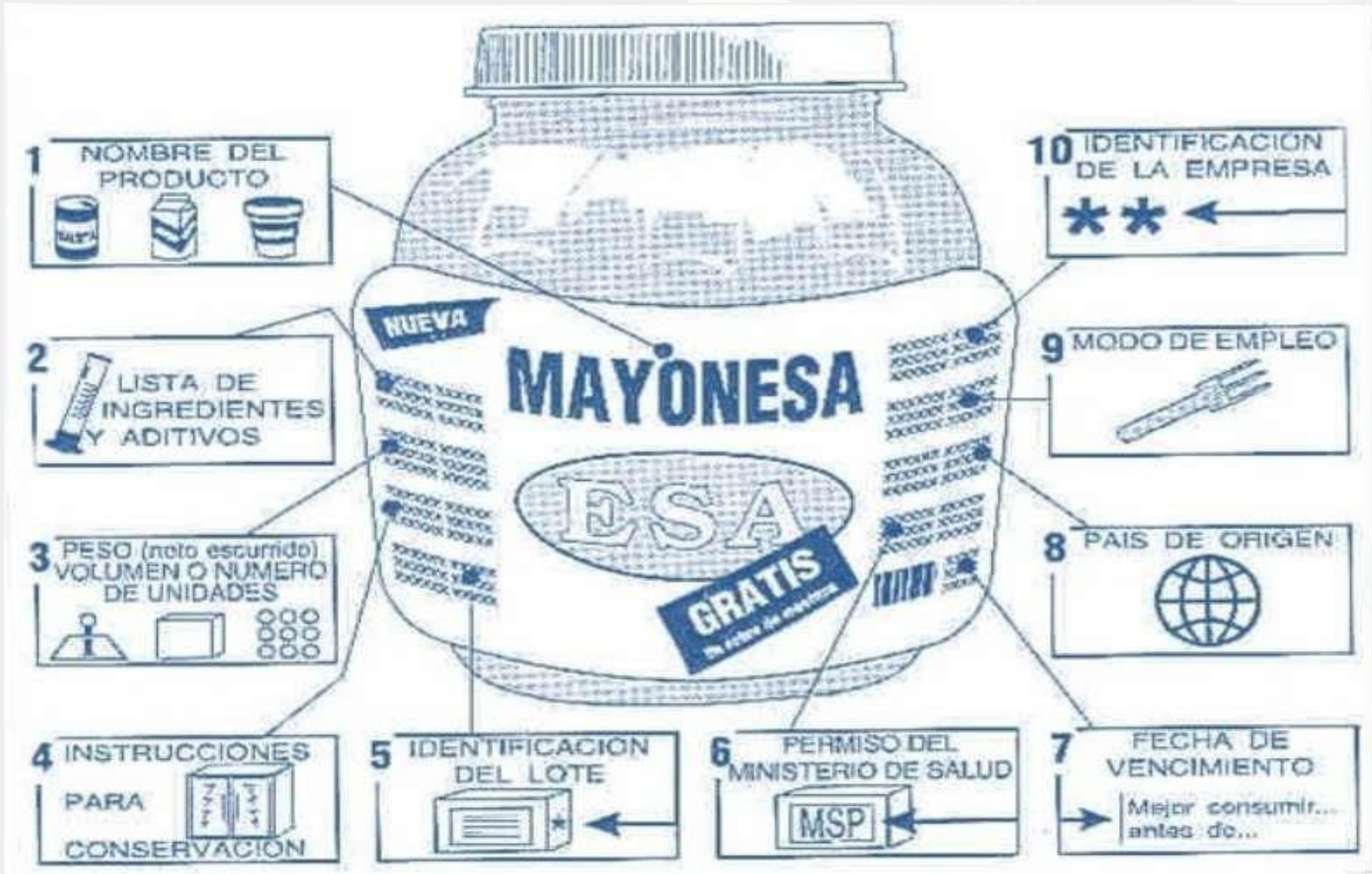
# Sin metadatos fiables, una organización no sabe qué datos tiene, qué representan los datos y cómo se transforman/mueven entre los sistemas, quién tiene acceso a ellos o qué calidad tienen sus productos de datos

*“With reliable Metadata an organization does not know what data it has, what the data represents and how it moves through the systems, who has access to it, or what it means for the data to be of high quality”. (fuente DmBok2)*

## ¿Qué es un Metadato?

La definición más común de metadatos, "datos sobre datos", es engañosamente simple. El tipo de información que puede clasificarse como Metadatos es muy amplio.

- **Proporcionan información** sobre procesos técnicos y comerciales, reglas y limitaciones de los datos y estructuras lógicas y físicas de los datos.
- **Describen** los datos propiamente dichos (por ejemplo, bases de datos, elementos de datos, modelos de datos), los conceptos que representan los datos (por ejemplo, procesos comerciales, sistemas de aplicación, código de software, infraestructura tecnológica) y las conexiones (relaciones) entre los datos y los conceptos.
- **Ayudan a una organización a comprender** sus datos, sus sistemas y sus flujos de trabajo.
- **Permiten evaluar la calidad de los datos** y son parte integrante de la gestión de las bases de datos y otras aplicaciones.
- **Contribuyen a la capacidad** de procesar, mantener, integrar, asegurar, auditar y gobernar otros datos.



## IMPORTANTE PARA EL EXAMEN

### Metadatos: Tipología de Metadatos

Los metadatos suelen clasificarse en tres tipos: de negocios, técnicos y operativos.

#### Metadatos de Negocio

- Los metadatos de negocios se centran en gran medida en el contenido y la condición de los datos e incluyen detalles relacionados con la gobernanza de los datos.
- Incluyen los nombres y definiciones no técnicos de conceptos, áreas temáticas, entidades y atributos; tipos de datos de atributos y otras propiedades de atributos; descripciones de rangos; cálculos; algoritmos y reglas de negocios; valores de dominio válidos y sus definiciones.

#### Metadatos Técnicos

- Los metadatos técnicos proporcionan información sobre los detalles técnicos de los datos, los sistemas que los almacenan y los procesos que los mueven dentro de los sistemas y entre ellos.

#### Metadatos Operacionales

- Metadatos operacionales describen los detalles del procesamiento y el acceso a los datos.

## Ejemplos de metadatos según su tipología:

| Metadatos de negocio                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Metadatos Técnicos                                                                                                                                                                                                                                                                                            | Metadatos Operacionales                                                                                                                                                                                                                                                                                                                 |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>- Definiciones y descripciones de conjuntos de datos, tablas y columnas</li> <li>- Modelos de datos</li> <li>- Reglas de negocio, transformación, cálculos y derivaciones</li> <li>- Reglas de calidad de datos y resultados de las mediciones</li> <li>- Estándares de datos</li> <li>- Nivel de seguridad/Privacidad de los datos</li> <li>- Problemas conocidos de los datos</li> <li>- Origen de datos y linaje</li> <li>- ...</li> </ul> | <ul style="list-style-type: none"> <li>- Nombres de columnas y tablas de datos físicas</li> <li>- Propiedades de la columna</li> <li>- Propiedades del objeto de base de datos</li> <li>- Permisos de acceso</li> <li>- Relaciones entre los modelos de datos y los activos físicos</li> <li>- ...</li> </ul> | <ul style="list-style-type: none"> <li>- Registros de errores</li> <li>- Patrones volumétricos y de uso</li> <li>- Reglas y acuerdos para compartir datos</li> <li>- Resultado de métricas de auditorías, balances y controles</li> <li>- Anomalías de horarios</li> <li>- Historia de extractos y resultados</li> <li>- ...</li> </ul> |

Fuera de la tecnología de la información, por ejemplo, en bibliotecas o ciencias de la información, los Metadatos se describen utilizando un conjunto diferente de categorías:

- **Los metadatos Descriptivos** (por ejemplo, el título, el autor y el tema) describen un recurso y permiten su identificación y recuperación.
- **Los Metadatos Estructurales** describen las relaciones dentro de los recursos y entre ellos, así como sus partes componentes (por ejemplo, número de páginas, número de capítulos).
- **Los Metadatos Administrativos** (por ejemplo, números de versión, fechas de archivo) se utilizan para gestionar los recursos a lo largo de su ciclo de vida

## Metadata Management

Si el metadato es la base del Gobierno de los Datos es absolutamente necesario gestionarlo, podemos hacerlo con una herramienta de Metadata Management o con un producto de Data Governance, una cosa es segura, tenemos que pasar por estas fases:

**Clasificación:** Los datos pueden tener muchas características por las que se pueden agrupar o clasificar. Tener esta categorización permitirá organizarlos y gestionarlos. Los datos pueden ser clasificados por cada una de las siguientes características:

- Tema - Ej: datos financieros, datos de estudiantes, datos de recaudación de fondos, datos de salud, datos de productos, etc.
- Uso - Ej: transaccional, analítico, regulatorio, etc.
- Actualización - Ej: datos en vivo y actuales, históricos, predictivos, etc.
- Contenido - Ej.: datos geoespaciales, datos de máquinas, datos estructurados vs. no estructurados, etc.
- Alcance - Ej: datos empresariales, externos, departamentales, maestros, etc.

**Descripción:** Describir los datos ayuda a entender tanto sus aspectos lógicos como físicos. Los datos descritos deben incluir:

- **Significado de los datos** - Definiciones comerciales, entidades de modelado de datos y atributos
- **Estructura de los datos** - Descripción de los objetos de datos (entidades, tablas, registros, etc.), sus agrupaciones y relaciones lógicas
- **Contenido de los datos** - Los tipos de datos como la fecha, la moneda, el texto, el número, etc.
- **Valores de los datos** - Qué valores están permitidos, qué datos de referencia están disponibles, qué patrones o rangos de valores deben seguir, qué restricciones deben cumplir, etc.
- **Linaje de los datos** -Cuál es la fuente de los datos, cómo se crearon, derivaron y/o calcularon los datos, cómo se transformaron, etc.

**Orientación:** El Metadato puede servir de guía a cualquier usuario técnico o empresarial para encontrar los datos que necesita, a través de motores de búsqueda u otros procesos. Estos metadatos de orientación pueden estar compuestos por

- **Palabras clave** - Estos podrían ser cualquier metadato descrito hasta ahora
- **Taxonomías** - Otro ejemplo más de cómo la clasificación ayuda
- **Sellos de fecha y hora** - Normalmente se añaden automáticamente a nivel de tabla o fila
- **Informes asociados, procesos, personas** - Saber dónde surgen los datos, quiénes son los usuarios o administradores de los datos, cómo se capturan y transforman los datos podría servir como un buen punto de partida para encontrar lo que se necesita
- **Sinónimos, alias, términos relacionados**

**Control:** Los metadatos pueden proporcionar los conocimientos necesarios para determinar qué controles deben aplicarse a los datos y qué datos deben ser controlados. Hace cumplir las limitaciones debidas a:

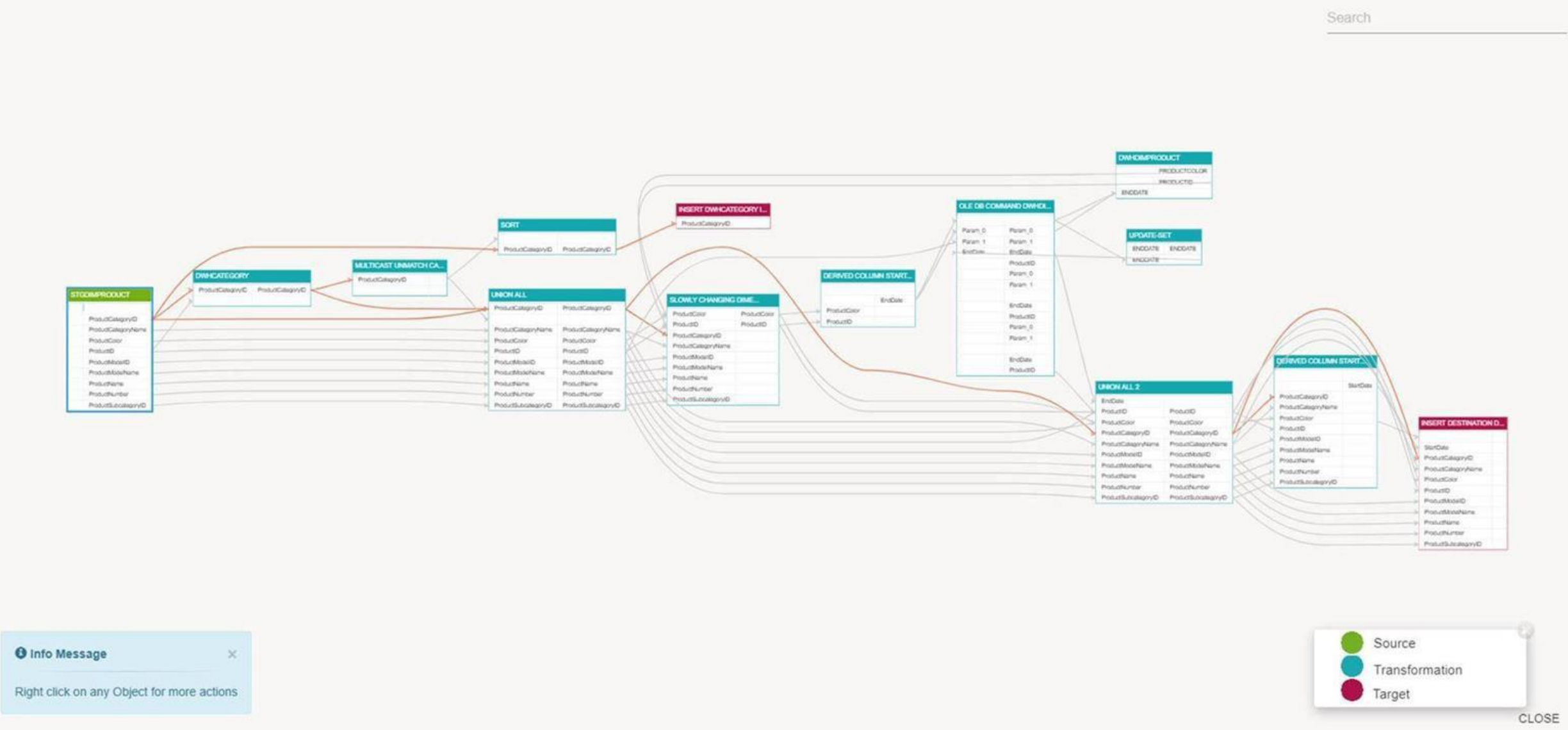
- Cumplimiento de la normativa y políticas internas
- Retención y archivo
- Privacidad y seguridad
- Niveles de servicio y requisitos comerciales
- Requisitos técnicos

## Técnicas: Data Lineage y Análisis de Impacto

Un beneficio clave de descubrir y documentar Metadatos sobre los activos físicos es proporcionar información sobre cómo se transforman los datos a medida que se mueven entre sistemas. Muchas herramientas de Metadatos contienen información sobre lo que está sucediendo a los datos dentro de sus entornos y proporcionan capacidades para ver el linaje a través de los sistemas o aplicaciones que interactúan.

- El alcance del linaje de datos determina el volumen de metadatos necesario para representar su linaje de datos.
- El linaje de datos proporciona una pista de auditoría de los puntos de datos en el nivel granular más alto, pero la presentación del linaje se puede hacer en varios niveles de profundidad para simplificar la información que se pone a disposición.

# Data Lineage



## Análisis de Impacto

El linaje documentado ayuda tanto a las personas de negocio como a los técnicos a utilizar los datos.

Permite:

- Reducir el tiempo de investigación de anomalías
- Analizar impactos por cambios potenciales o resultados desconocidos
- Comprender el proceso de ingesta, así como reportes y análisis de usuarios finales

Los informes de impacto describen qué componentes se ven afectados por un posible cambio agilizando y racionalizando las tareas de estimación y mantenimiento.

Sin metadatos confiables, una organización no sabe qué datos tiene, qué representan los datos y cómo se mueven a través de los sistemas, quién tiene acceso a ellos o qué significa que los datos sean de alta calidad.



A. Verdadero

B. Falso

---

---

---

## Business Glossary

El propósito de un glosario de negocios (business glossary) es documentar y almacenar los conceptos y la terminología comercial de una organización, así como las definiciones y las relaciones entre esos términos.

La aplicación del glosario de negocios está estructurada para satisfacer los requisitos funcionales de las tres audiencias principales:

- **Usuarios de negocios:** Los analistas de datos, los analistas de investigación, administración y el personal ejecutivo utilizan el glosario empresarial para comprender la terminología y los datos.
- **Data Steward:** los *Data Stewards* utilizan el glosario empresarial para gestionar el ciclo de vida de los términos y definiciones y para mejorar el conocimiento de la empresa asociando los activos de datos con los términos del glosario; por ejemplo, vinculando los términos a las métricas empresariales, los informes, el análisis de la calidad de los datos o los componentes tecnológicos. Los administradores de datos plantean cuestiones de terminología y uso y ayudan a resolver las diferencias en toda la organización.
- **Usuarios técnicos:** Los usuarios técnicos utilizan el glosario empresarial para tomar decisiones sobre arquitectura, diseño de sistemas y desarrollo, y para realizar análisis de impacto.

- **Un Business Glossary se centra en el lenguaje de los negocios y es fácil de entender en cualquier entorno de business, desde salas de juntas hasta equipos de tecnología.**
- Los términos empresariales no están destinados a definir datos, metadatos, transformaciones o ubicaciones, sino más bien a **definir lo que cada término significa en un sentido empresarial**. ¿Qué queremos decir con una conversión? ¿Una venta? Este tipo de preguntas se pueden responder con un glosario de negocios.
- Disponer de un glosario de negocios aporta una **comprensión común** del vocabulario utilizado en toda una organización.
- El ámbito de aplicación de un glosario de negocios debe ser de toda la empresa o, al menos, toda la división en los casos en que las diferentes divisiones tengan una terminología de negocios significativamente diferente.
- Debido a su alcance y a los conocimientos técnicos necesarios, **la responsabilidad del glosario empresarial recae en toda la empresa y no solamente en tecnología.**

## Data Dictionary (Diccionario de datos)

Un diccionario de datos debe centrarse en las descripciones y detalles que conlleva el almacenamiento de datos está directamente ligado a los metadatos.

- Debería haber un diccionario de datos para **cada fuente de datos** de la empresa.
- El diccionario de datos incluye detalles sobre los datos como el tipo de datos, la longitud permitida, el linaje, las transformaciones, etc.
- Estos metadatos ayudan a los **arquitectos, ingenieros y científicos de datos** a entender cómo unir, consultar y reportar los datos, y también explica la granularidad.
- Debido a la necesidad de conocimientos técnicos y metadatos, la responsabilidad de la propiedad de un diccionario de datos recae en la tecnología, frecuentemente con funciones como administradores de bases de datos, ingenieros de datos, arquitectos de datos y/o administradores de datos.

## Data Catalog (Data Marketplace)

El catálogo de datos sirve como un directorio de un solo punto para localizar información y además proporciona el mapeo entre el glosario de negocios y los diccionarios de datos.

- **El catálogo de datos es un activo de toda la empresa** que proporciona una única fuente de referencia para la localización de cualquier conjunto de datos necesarios para diversas necesidades como las operacionales, BI, analíticas, de ciencia de los datos, etc.
- Si una división de una empresa es significativamente diferente de otras, sería razonable que el catálogo de datos fuera exclusivo de la división y no de la empresa.
- Normalmente el catálogo de datos se desarrolla después de la creación satisfactoria tanto del Business Glossary como de los diccionarios de datos, pero también puede ensamblarse de forma incremental a medida que los otros dos activos evolucionen con el tiempo.

## Definición de Data Catalog

David Stodder, Director de Investigación de TDWI, definió el catálogo de datos como:

*"Una especie de piedra de Rosetta que permite a los usuarios, desarrolladores y administradores encontrar y aprender sobre los datos - y para los profesionales de la información para organizar, integrar y curar adecuadamente los datos para los usuarios. "*

## Data Catalog el nuevo reto del Data Management y de la Analitica (según Gartner)

La demanda de catálogos de datos se dispara a medida que las organizaciones luchan por inventariar los datos distribuidos activos para facilitar la monetización de los datos y ajustarse a los reglamentos. Los proyectos de catálogos de datos caerán ...si los líderes de datos y análisis no los vinculan con datos más amplios... necesidades de gestión.

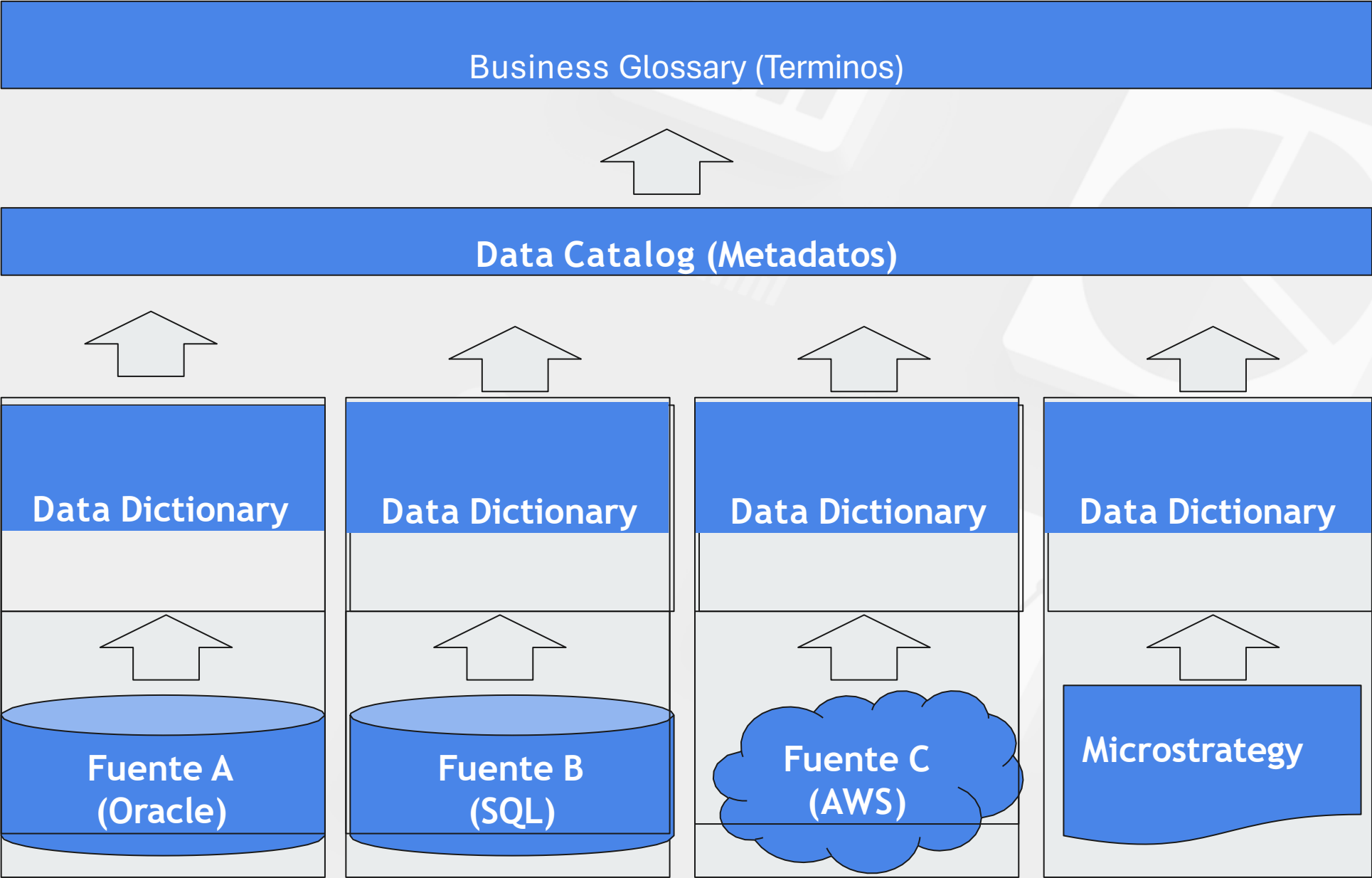
### Impactos

- Los catálogos de datos permiten a los líderes en materia de datos y análisis introducir un gobierno ágil de la información, y gestionar la dispersión de datos y las cadenas de suministro de información, en apoyo de las iniciativas comerciales digitales.
- Se están adoptando catálogos de datos en un esfuerzo por remediar la falta de disciplina en los metadatos pero la profusión de ofertas y enfoques de los proveedores, y la publicidad, crean incertidumbre sobre el valor y la selección de una solución óptima de catálogo de datos.
- Abrazar los catálogos de datos sin un plan de gobierno de la información que los vincule a la más amplia Las necesidades de gestión de metadatos dejarán a los líderes de datos y análisis incapaces de gestionar eficazmente y monetizar sus activos de datos a largo plazo.

## Recomendaciones

Para los líderes de datos y análisis que evalúan los catálogos de datos como parte de una estrategia de gestión de datos:

- **Utilizar los catálogos de datos para mejorar el inventario de los activos de información distribuidos disponibles y para cartografiar cadenas de suministro de información, convirtiéndolas en un componente esencial de su gestión de datos estrategia.**
- Proceder con el conocimiento de que los catálogos de datos incrustados específicos de la herramienta (como los que se entregan como parte de una distribución de Hadoop, un Data Lake en nube o una herramienta de preparación de datos) mejorará los datos de uso, confianza y compartimentación **sólo en el contexto de esa herramienta en particular.**
- Desplegar catálogos de datos con la capacidad de escalar más allá de un caso de uso estrecho (o táctico) necesidades de datos y análisis estratégicos, asegurándose de que son conectados con la gestión de metadatos y los programas de gobernanza de la información de la empresa en general.



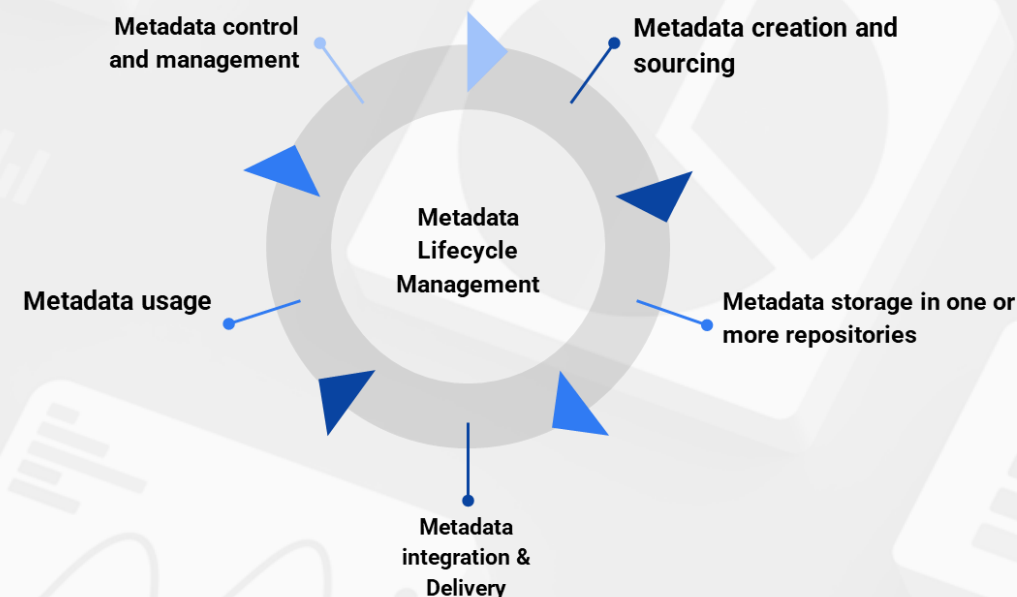
## Herramientas de gestión de configuración

- Las **CMDB** (herramientas o bases de datos de gestión de la configuración) proporcionan la capacidad de gestionar y mantener metadatos específicamente relacionados con los activos de TI, las relaciones entre ellos y los detalles contractuales de los activos.
- Cada elemento de la base de datos CMDB se denomina elemento de configuración (CI = Configuration Item). Los Metadatos estándares se recopilan y administran por cada tipo de CI.
- Los repositorios proporcionan mecanismos para vincular los activos en el repositorio de Metadatos con los detalle físicos reales de la implementación en CMDB para dar una imagen completa de los datos y las plataformas.

## IMPORTANTE PARA EL EXAMEN

Todas las soluciones de gestión de metadatos incluyen capas arquitectónicas que incluyen:

- Metadata creation and sourcing
- Metadata storage in one or more repositories
- Metadata integration
- Metadata delivery
- Metadata usage
- Metadata control and management



***Un repositorio centralizado de metadatos permite tener alta disponibilidad, ya que es independiente de los sistemas de origen.***

## IMPORTANTE PARA EL EXAMEN

### ¿Cuáles son los entregables del Metadata Management Context Diagram?

- Metadata Strategy
- Metadata Standards
- Data Lineage
- Metadata architecture

***Los metadatos son clave tanto para datos estructurados como para no estructurados***

# El acrónimo C MDB significa:

- A. Herramientas o bases de datos de gestión de personalización
- B. Tecnologías o bases de datos de gestión clásica
- C. Tecnologías o bases de datos de gestión en caché
-  D. Herramientas o bases de datos de gestión de configuración
- E. Herramientas o bases de datos de gestión en caché
- F. Tecnologías o bases de datos de monitoreo clásico



# Construir un Data Catalog sin morir en el intento

# Cómo construir un Data Catalog / Data Marketplace



## 1. Recogida de metadatos:

La recopilación de metadatos de todos los datos presentes es el primer paso en la construcción de un catálogo de datos.

Un catálogo de datos rastrea todos los datos presentes en el repositorio y copia los metadatos en el catálogo. Estos metadatos se utilizan para identificar conjuntos de datos, tablas de datos y archivos.



# Cómo construir un Data Catalog / Data Marketplace

## 2. Construcción de un diccionario de datos:

Una vez recogidos los metadatos, el siguiente paso es **construir un diccionario de datos que contenga la descripción de todos los metadatos**. Existen múltiples programas informáticos que facilitan la creación de un diccionario de datos o se puede utilizar un simple Excel para construirlo. Lo ideal sería alinearlos ya a los términos del **Business Glossary**

# Cómo construir un Data Catalog / Data Marketplace



## 2. Perfilar los datos:

El siguiente paso en la creación de un catálogo de datos es **perfilar los datos disponibles**, lo que ayuda a las partes interesadas a visualizar y comprender los datos. En esta fase es necesario involucrar tanto los **Data Owners** que los **Data Stewards**

# Cómo construir un Data Catalog / Data Marketplace

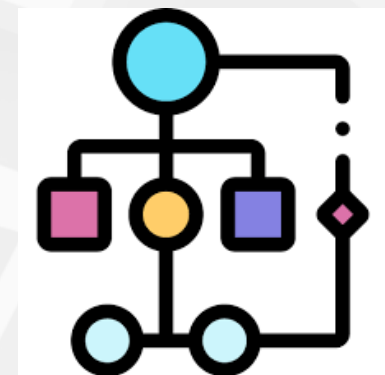


## 3. Marcar la relación entre los datos:

Marcar la relación entre los conjuntos de datos es fundamental, ya que ayuda a las partes interesadas a **entender fácilmente la relación de los datos entre las fuentes de datos.**

Esta relación puede ser marcada manualmente por individuos o por sistemas avanzados, que aprenden de las consultas realizadas entre conjuntos de datos o valores similares. (ojo con los automatismos, siempre es mejor controlar sobre todo si los datos son sensibles y valioso para las analisis)

# Cómo construir un Data Catalog / Data Marketplace



## 4 .Construcción del linaje:

Un linaje representado visualmente es muy útil para **establecer la trayectoria de los datos**. Desde el origen hasta el destino. Los diferentes procesos que intervienen en el flujo de datos se explican a través del linaje. Esto permite a las partes interesadas encontrar la causa raíz de un error muy fácilmente con sólo seguir el linaje.

El linaje puede ser horizontal o vertical.

La lectura inversa del linaje da pistas de auditoría y se llama Análisis de Impacto

# Cómo construir un Data Catalog / Data Marketplace



## 5 .Organización de los datos:

Los datos dentro de un archivo o una tabla se presentan de forma técnica. Esto puede o no tener sentido según los requisitos del negocio. **Es necesario realizar un esfuerzo manual para organizar los datos de manera que puedan ser fácilmente comprensibles y fiables para los usuarios de la empresa.**

Algunas de las formas de hacerlo son el **etiquetado de los datos**, la **organización de los datos en función del uso** y la función del usuario y la automatización.

# Cómo construir un Data Catalog / Data Marketplace



## 6 .Asegurar y trazar el acceso:

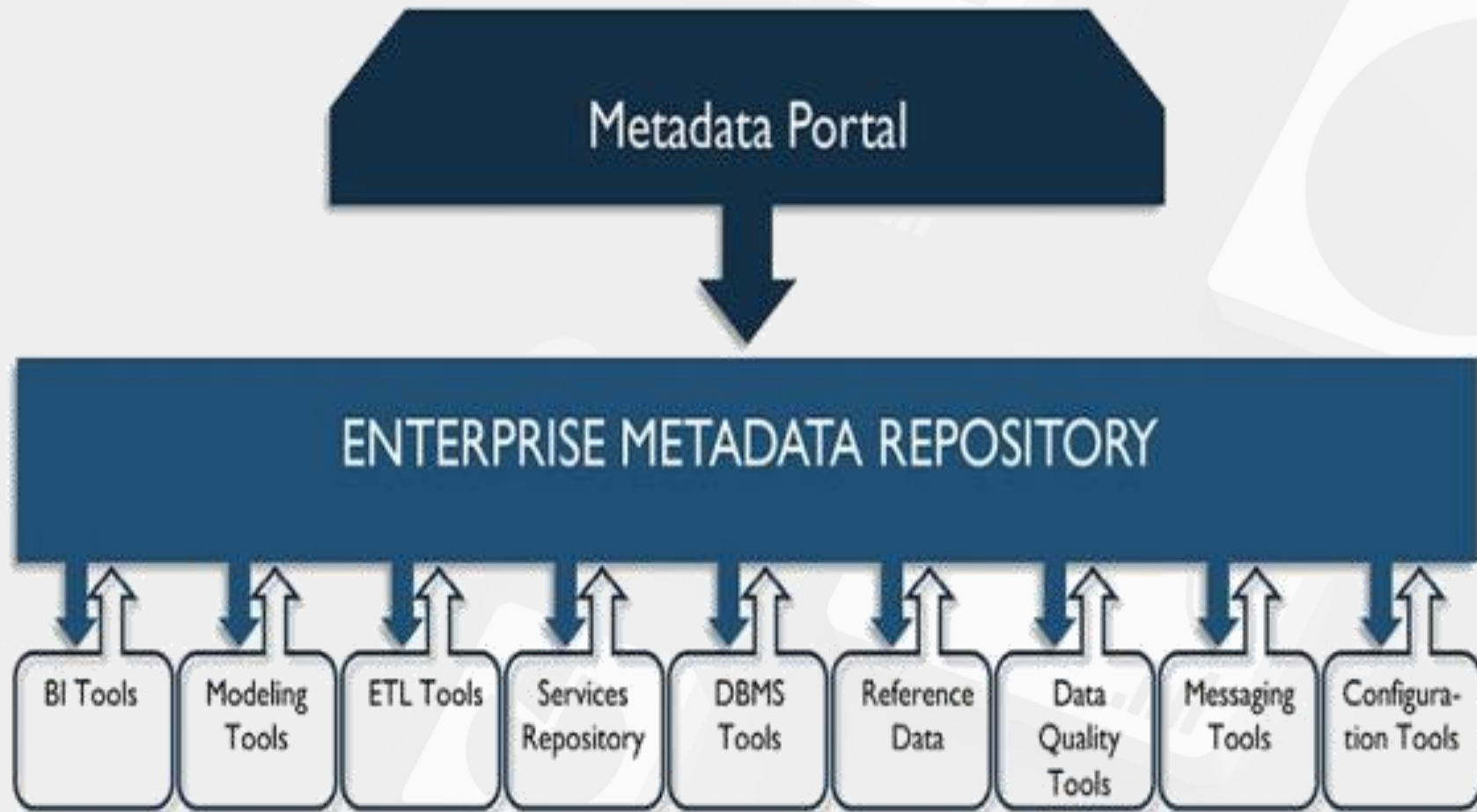
Para un mayor uso del catálogo de datos, éste debe ser fácilmente accesible dentro de la pila de datos.

Puede estar en un portal interno o en alguna herramienta siempre que toda la seguridad de acceso sea activa. Allí van los datos más valiosos de la empresa, no todo el mundo puede acceder a ellos y quien lo hace tiene que ser trazado.



# Metadata Lifecycle y repositorios de metadatos

# Centralized Metadata Architecture





# Arquitectura de metadatos centralizada

Una arquitectura centralizada consiste en un único repositorio de metadatos que contiene copias de los metadatos de las distintas fuentes.

Las organizaciones que buscan un alto grado de consistencia dentro del repositorio común de metadatos pueden beneficiarse de una arquitectura centralizada. Las ventajas de un repositorio centralizado incluyen:.

- **Alta disponibilidad**, ya que es independiente de los sistemas de origen
- **Rápida recuperación de metadatos**, ya que el repositorio y la consulta residen juntos
- **Estructuras de bases de datos resueltas** que no se ven afectadas por la naturaleza propietaria de sistemas comerciales o de terceros
- Los metadatos extraídos pueden transformarse, personalizarse o mejorarse con Metadatos adicionales que pueden no residir en el sistema de origen, mejorando la calidad

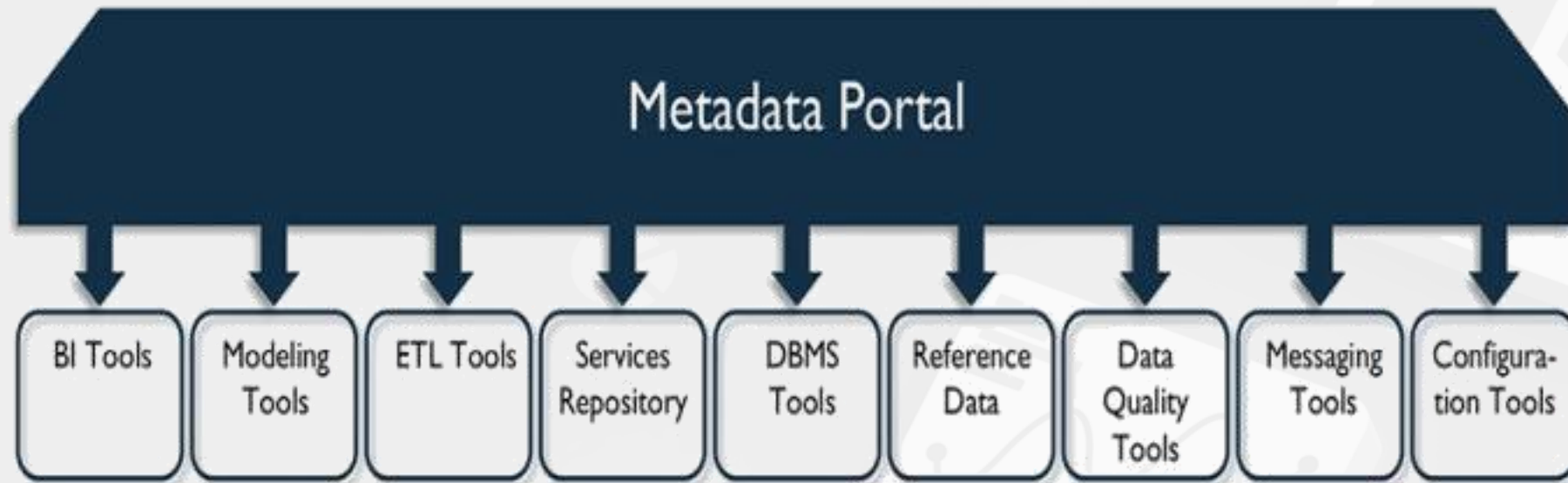


# Arquitectura de metadatos centralizada

- Son necesarios procesos complejos para garantizar que los cambios en la fuente
- Los metadatos se replican rápidamente en el repositorio
- El mantenimiento de un repositorio centralizado puede ser costoso
- La extracción puede requerir módulos personalizados o **middleware**
- La validación y el mantenimiento del código personalizado pueden aumentar las exigencias tanto al personal informático interno como a los proveedores de software



# Distributed Metadata Architecture



# Arquitectura de metadatos distribuida

Una arquitectura completamente distribuida mantiene un único punto de acceso. El motor de recuperación de metadatos responde a las solicitudes de los usuarios recuperando los datos de los sistemas en tiempo real; no hay un repositorio persistente.

En esta arquitectura, el entorno de gestión de metadatos mantiene los catálogos de los sistemas fuente y la información de búsqueda necesaria para procesar eficazmente las consultas y búsquedas de los usuarios.

# Arquitectura de metadatos distribuida

- Los metadatos siempre son actuales y válidos como sea posible porque se recuperan de su fuente
- Las consultas se distribuyen, lo que puede **mejorar el tiempo de respuesta y procesamiento**
- Las solicitudes de metadatos de los sistemas propietarios **se limitan a la consulta**, en lugar de requerir un conocimiento detallado de las estructuras de datos propietarias, por lo que se minimiza el esfuerzo de implementación y mantenimiento requerido
- El desarrollo del procesamiento automatizado de consultas de metadatos es probablemente más sencillo, requiere una intervención manual mínima
- **Se reduce el procesamiento por lotes**, sin procesos de replicación o sincronización de metadatos. procesos de sincronización



# Arquitectura de metadatos distribuida

- No hay capacidad para **admitir entradas de metadatos definidas por el usuario** o insertadas manualmente, ya que no hay un repositorio en el que colocar estas adiciones
- Estandarización de la presentación de metadatos de varios sistemas
- Las capacidades de consulta se ven **directamente afectadas por la disponibilidad de los sistemas fuente participantes**
- La calidad de los metadatos **depende exclusivamente de los sistemas fuente participantes.**

# Arquitectura híbrida de metadatos

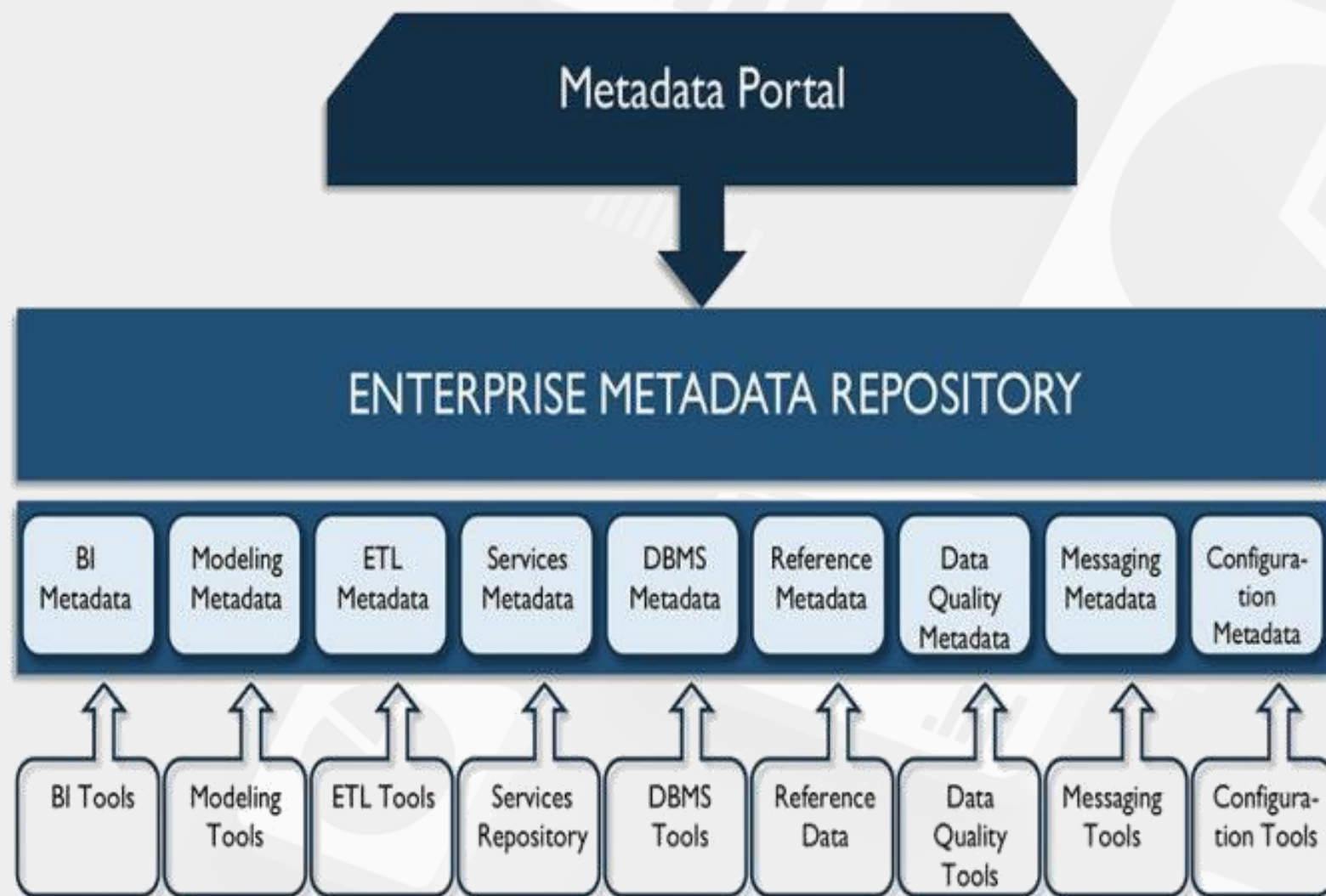
Una arquitectura híbrida combina las características de las arquitecturas centralizadas y distribuidas.

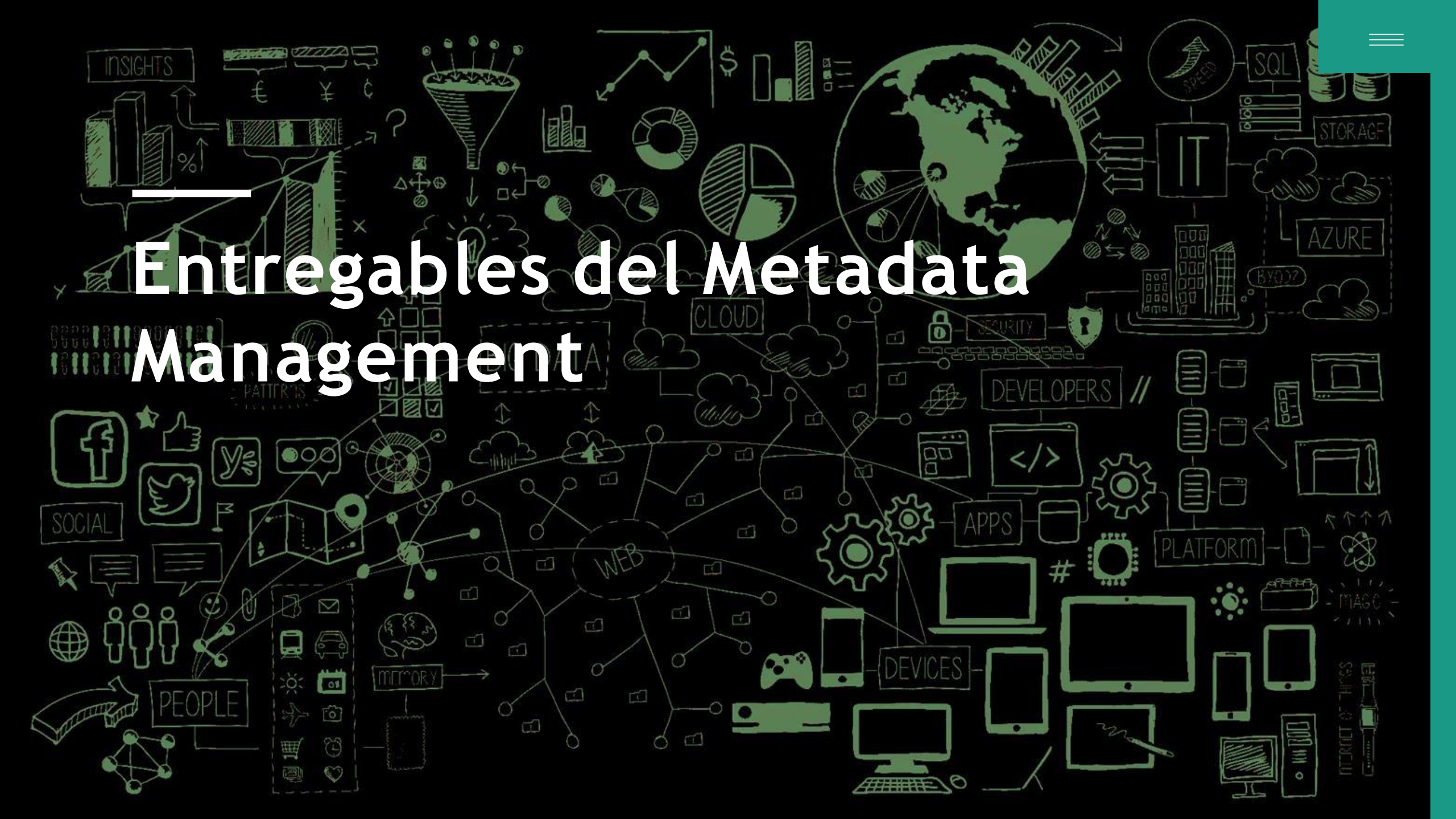
Los metadatos siguen pasando directamente de los sistemas de origen a un repositorio centralizado. Sin embargo, el diseño del repositorio sólo tiene en cuenta los metadatos añadidos por el usuario, los elementos críticos estandarizados

La arquitectura se beneficia de la recuperación casi en tiempo real de los metadatos desde su fuente y de la mejora de los metadatos para satisfacer las necesidades del usuario de la manera más eficaz, cuando sea necesario.

El enfoque híbrido reduce el esfuerzo de la intervención manual de TI y la funcionalidad de acceso codificada a medida de los sistemas propietarios. sustancial de los Metadatos y de las fuentes de Metadatos. fuentes.

# Hybrid Metadata Architecture





# Entregables del Metadata Management

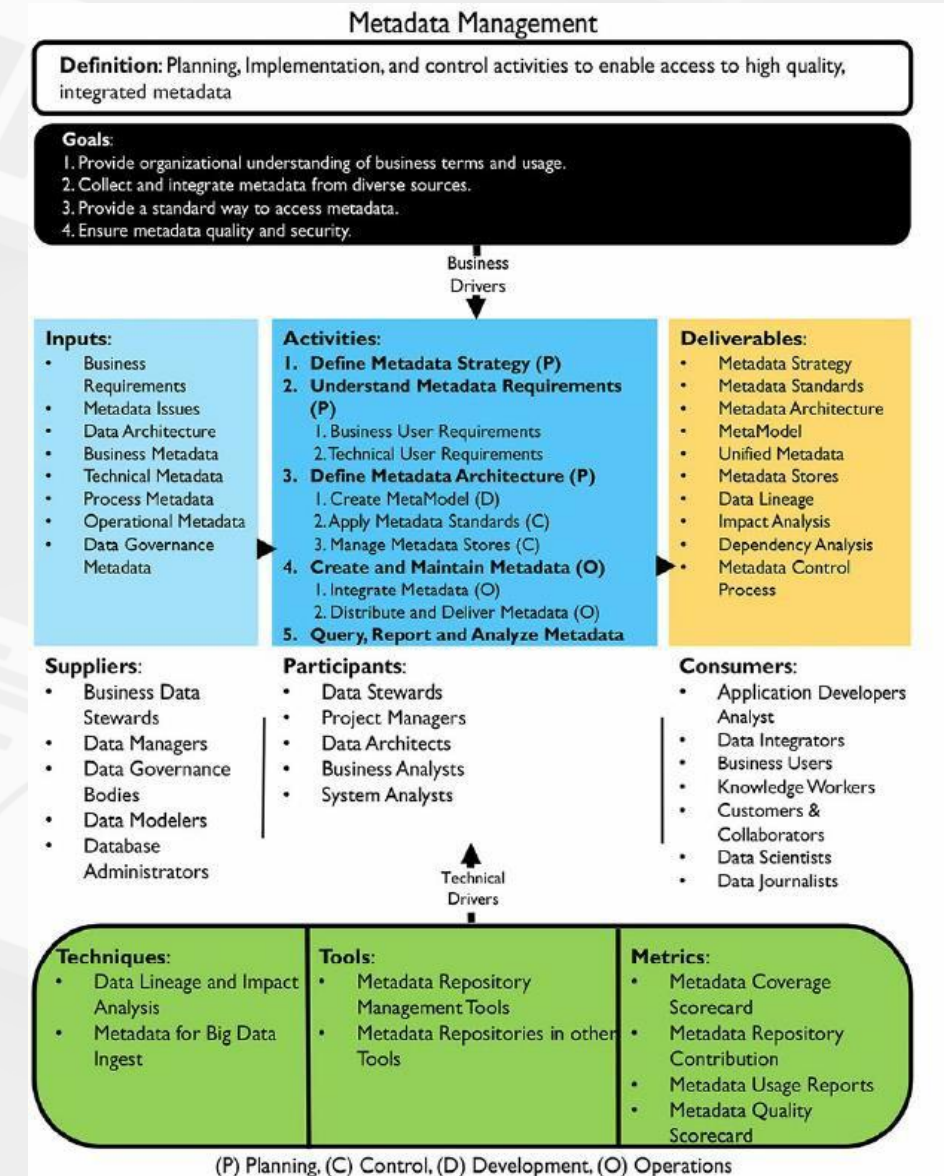
# IMPORTANTE PARA EL EXAMEN

## ¿Cuales son los entregables del Metadata Management Context Diagram?

Los entregables en el diagrama de contexto de la gestión de metadatos incluyen:

- Metadata Strategy
- Metadata Standards
- Metadata Architecture
- MetaModel
- Unified Metadata
- Metadata Stores
- Data Lineage
- Impact Analysis
- Dependency Analysis
- Metadata Control Process

Metadata is essential to the management of unstructured data as it is to the management of structured data.



# Los entregables en el diagrama de contexto de Gestión de Metadatos incluyen:

- ✓ A. Estrategia de metadatos
- ✓ B. Estándares de metadatos
- ✓ C. Linaje de datos
- ✓ D. Arquitectura de metadatos
- ✓ E. Diseño de metadatos
- F. Almacenamiento y operaciones de datos

# Los objetivos de la gestión de metadatos incluyen:

- A. Gestionar el rendimiento de los activos de datos
-  B. Asegurar la calidad, consistencia, actualización y seguridad de los metadatos
- C. Gestionar la disponibilidad de los datos a lo largo del ciclo de vida de los datos
-  D. Proveer formas estandarizadas de hacer accesibles los metadatos a los consumidores de metadatos
- E. Gestionar el rendimiento de las transacciones de datos
-  F. Establecer o hacer cumplir el uso de estándares de metadatos técnicos para permitir el intercambio de datos

# Algunas herramientas...



# Metadata Tools

## Microsoft Azure Data Catalog (Pureview)

Azure Data Catalog es un servicio de nubes totalmente gestionado. Permite a los usuarios descubrir las fuentes de datos que necesitan y entender las fuentes de datos que encuentran. Al mismo tiempo, Data Catalog ayuda a las organizaciones a obtener más valor de sus inversiones existentes.

- **Los datos permanecen en su ubicación actual**, pero se añade una copia de sus metadatos al Catálogo de Datos, junto con una referencia a la ubicación de la fuente de datos.
- **Los metadatos también se indexan para que cada fuente de datos sea fácilmente localizable** mediante la búsqueda y comprensible para los usuarios que la descubren.
- Una vez registrada una fuente de datos, sus metadatos pueden ser **enriquecidos**.
- Los metadatos pueden ser añadidos por el usuario que los registró o por otros usuarios de la empresa.

tags:product

X Q

Results Per Page: 10

Sort By: Relevance

Highlight

Open In ...

Delete

Searches

▼ Search Database

Product tag search

Share within the company.

Save Save As Clear All

Container: RLStest

Search Term: tags:product

▼ Saved Searches (1)

Product tag search

▼ Filters

User Tags

azure sql (5)

product (5)

sales (5)

Object Type

Table (5)

Source Type

SQL Server (5)

← Back to catalog

RLStest

Experts: azure sql product sales

5 Table

5 search results, 1 selected | ☐ Select All

|                                     |  | NAME                          | DESCRIPTION          | EXPERTS | TAGS               | CONTAINED IN | SOURCE TYPE | OBJECT TYPE | LAST UPDAT...       | LAST REGISTE...       |
|-------------------------------------|--|-------------------------------|----------------------|---------|--------------------|--------------|-------------|-------------|---------------------|-----------------------|
| <input type="checkbox"/>            |  | ProductModelProductDescrip... |                      |         | azure sql produ... | RLStest      | SQL Server  | Table       | 4/7/2019 10:34:3... | 4/7/2019 10:34:39 ... |
| <input type="checkbox"/>            |  | Product Model                 | Model of the product |         | azure sql produ... | RLStest      | SQL Server  | Table       | 4/8/2019 10:22:5... | 4/7/2019 10:34:39 ... |
| <input type="checkbox"/>            |  | ProductCategory               |                      |         | azure sql produ... | RLStest      | SQL Server  | Table       | 4/7/2019 10:34:3... | 4/7/2019 10:34:38 ... |
| <input type="checkbox"/>            |  | ProductDescription            |                      |         | azure sql produ... | RLStest      | SQL Server  | Table       | 4/7/2019 10:34:4... | 4/7/2019 10:34:40 ... |
| <input checked="" type="checkbox"/> |  | Product                       |                      |         | azure sql produ... | RLStest      | SQL Server  | Table       | 4/7/2019 10:34:4... | 4/7/2019 10:34:41 ... |

Excel (Top 1000)

Excel

SQL Server Data Tools

Power BI Desktop

Power Query

## Anjana Data

Anjana Data es una solución moderna de gestión de metadatos y gobierno de datos, diseñada para centralizar, gestionar y explotar el conocimiento de los datos de forma colaborativa orientada a gestionar los activos de datos.

- Estructura modular y escalable que se basa en un modelo de metadatos extensible, lo que permite adaptar la solución a las necesidades específicas de cada organización
- Permite configuración de políticas de gobierno y gestión de datos
- Incluye una variedad de APIs abiertas para facilitar la integración con otras herramientas de datos y ofrece conectores para cosechar y actualizar automáticamente metadatos, así como gestionar versiones y realizar ingeniería inversa de procesos de transformación y modelos de datos.

OpenMetaData

Analytics

GitBook

Calendar

Orbit

Orbit Info

TweetDeck

Calendly

HubSpot

Uppbeat

Upwork

QRCode Monkey

Community Stats...

reverse dictionary

Other Bookmarks

Open Metadata

Explore

Settings

Search for Table, Topics, Dashboards, Pipeline and ML Models

🔔

🚀

📖

?

M

sample\_data > ecommerce\_db > shopify > dim\_address

No Owner | No Tier | Type: Regular | Usage - 0th pctile | 5 queries | 12 columns | 725 rows

+ Add tag

Schema

Activity Feed & Task 0

Sample Data

Queries

Profiler

Data Quality

Lineage

Custom Properties

Manage

This dimension table contains the billing and shipping addresses of customers. You can join this table with the sales table to generate lists of the billing and shipping addresses. Customers can enter their addresses more than once, so the same address can appear in more than one row in this table. This table contains one row per customer address.

🔗

?

💬

Frequently Joined Tables

No information about joined tables.

Find in table...

| Name         | Type    | Data Quality | Description                                                                                              | Tags   |
|--------------|---------|--------------|----------------------------------------------------------------------------------------------------------|--------|
| 🔗 address_id | numeric | --           | Unique identifier for the address.                                                                       | + Tags |
| 🔗 shop_id    | numeric | --           | The ID of the store. This column is a foreign key reference to the shop_id column in the dim_shop table. | + Tags |
| first_name   | varchar | --           | First name of the customer.                                                                              | + Tags |
| last_name    | varchar | --           | Last name of the customer.                                                                               | + Tags |
| address1     | varchar | --           | The first address line. For example, 150 Elgin St.                                                       | + Tags |
| address2     | varchar | --           | The second address line. For example, Suite 800.                                                         | + Tags |

## Erwin Data Intelligence

La solución de administración de metadatos de erwin, el erwin Data Intelligence Suite (erwin DI), combina la administración de datos y los procesos de gobierno de datos en un flujo automatizado.



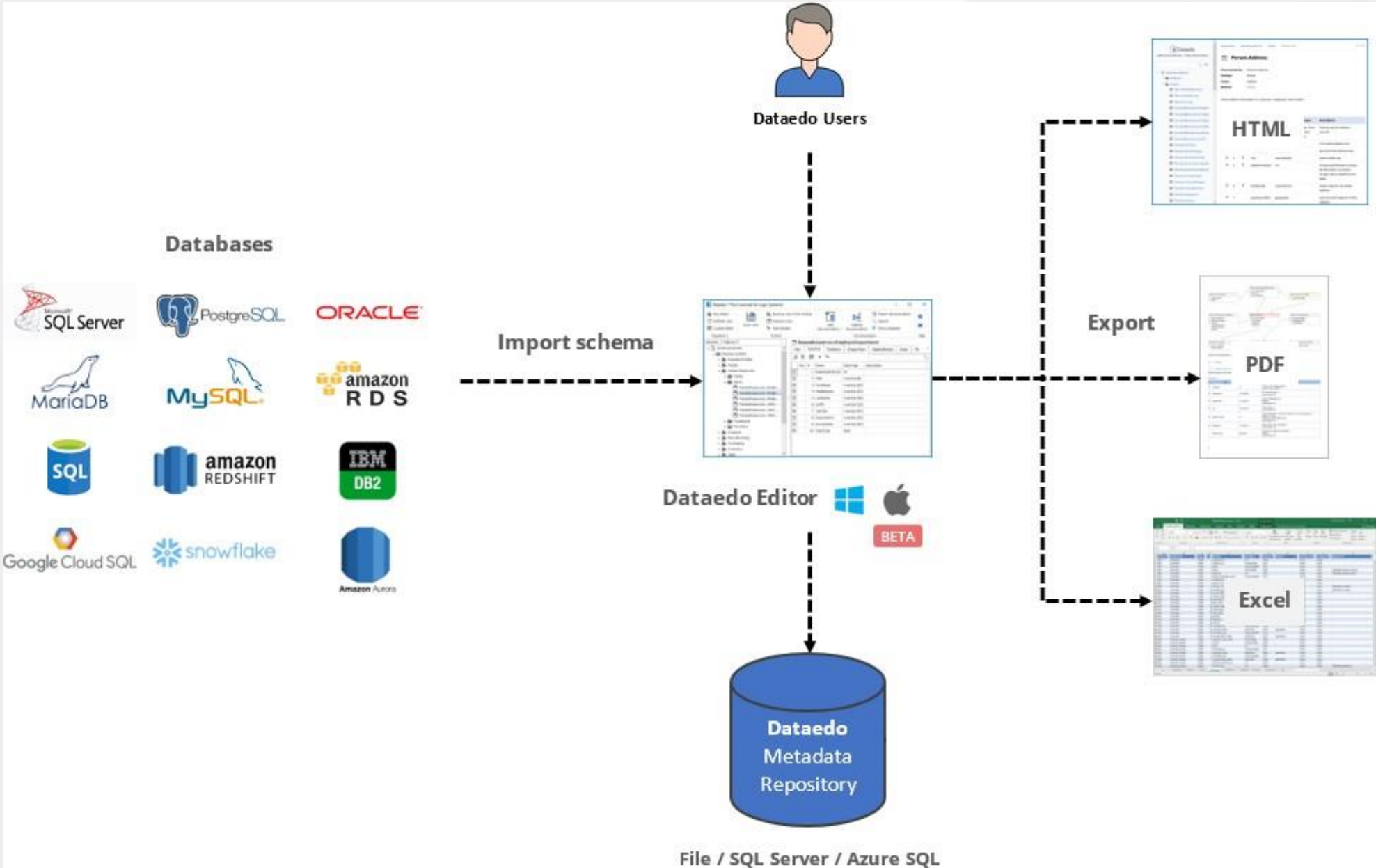
- Tiene la capacidad para cosechar, transformar y **alimentar automáticamente** los metadatos de una amplia gama de fuentes de datos, procesos operativos, aplicaciones empresariales y modelos de datos en un catálogo central de datos y luego hacerlos accesibles y comprensibles en el contexto de las vistas basadas en roles.
- Se sienta en un metamodelo común que es abierto, extensible y viene con un juego completo de **APIs**.
- Presenta una lista completa de **conectores** de datos estándar propiedad de erwin para la recolección automatizada, la actualización y la gestión de metadatos de versión controlada.
- Los conectores de datos inteligentes erwin opcionales permiten realizar **ingeniería inversa de todo tipo de códigos ETL** y se conectan bidireccionalmente con herramientas de informes y otras herramientas de ecosistemas.

## Dataedo

Dataedo es una poderosa herramienta de documentación de la base de datos. Permite crear diccionarios de datos (descripción de todas las tablas, columnas y relaciones), crear diagramas ER, documentar scripts de servidor (procedimientos almacenados, funciones y disparadores).



Es una aplicación de escritorio que se conecta a las bases de datos existentes e importa el esquema, y luego permite a los usuarios proporcionar descripciones y exportar la documentación a PDF, HTML o Excel.



## **Alation Data Catalog: Automatización del inventario de metadatos**

Alation utiliza técnicas de búsqueda perfeccionadas para simplificar y automatizar la creación de un inventario de los activos de datos.

- Alation se conecta a las fuentes de datos y el sistema rastrea e indexa los activos de datos almacenados en diferentes repositorios físicos, incluyendo bases de datos, archivos Hadoop y herramientas de visualización de datos.
- Alation ingiere automáticamente metadatos técnicos, permisos de usuario y descripciones de negocio en un repositorio central que es un recurso fundamental para todos los usuarios de datos de la organización.

Alation trabaja con estas fuentes de datos

- Oracle
- MySQL
- SQLServer
- Teradata
- Redshift
- Tableau

## IBM Watson® Knowledge Catalog

IBM Watson® Knowledge Catalog es un catálogo de datos abierto e inteligente para la gestión de datos empresariales y el gobierno, la calidad y la colaboración de los modelos de IA.\*

IBM Watson Knowledge Catalog ayuda a acelerar el valor comercial y los proyectos de IA como parte de una práctica de **DataOps** al proporcionar:

- Soporte de virtualización de datos en tiempo real
- Descubrimiento de datos y generación de metadatos automatizados
- ML-extraído el glosario de negocios de los términos regulatorios más comunes
- Enmascaramiento dinámico de datos para proteger los datos sensibles
- Escaneo automatizado y evaluaciones de riesgo de datos no estructurados a través de Watson™ Knowledge Catalog InstaScan

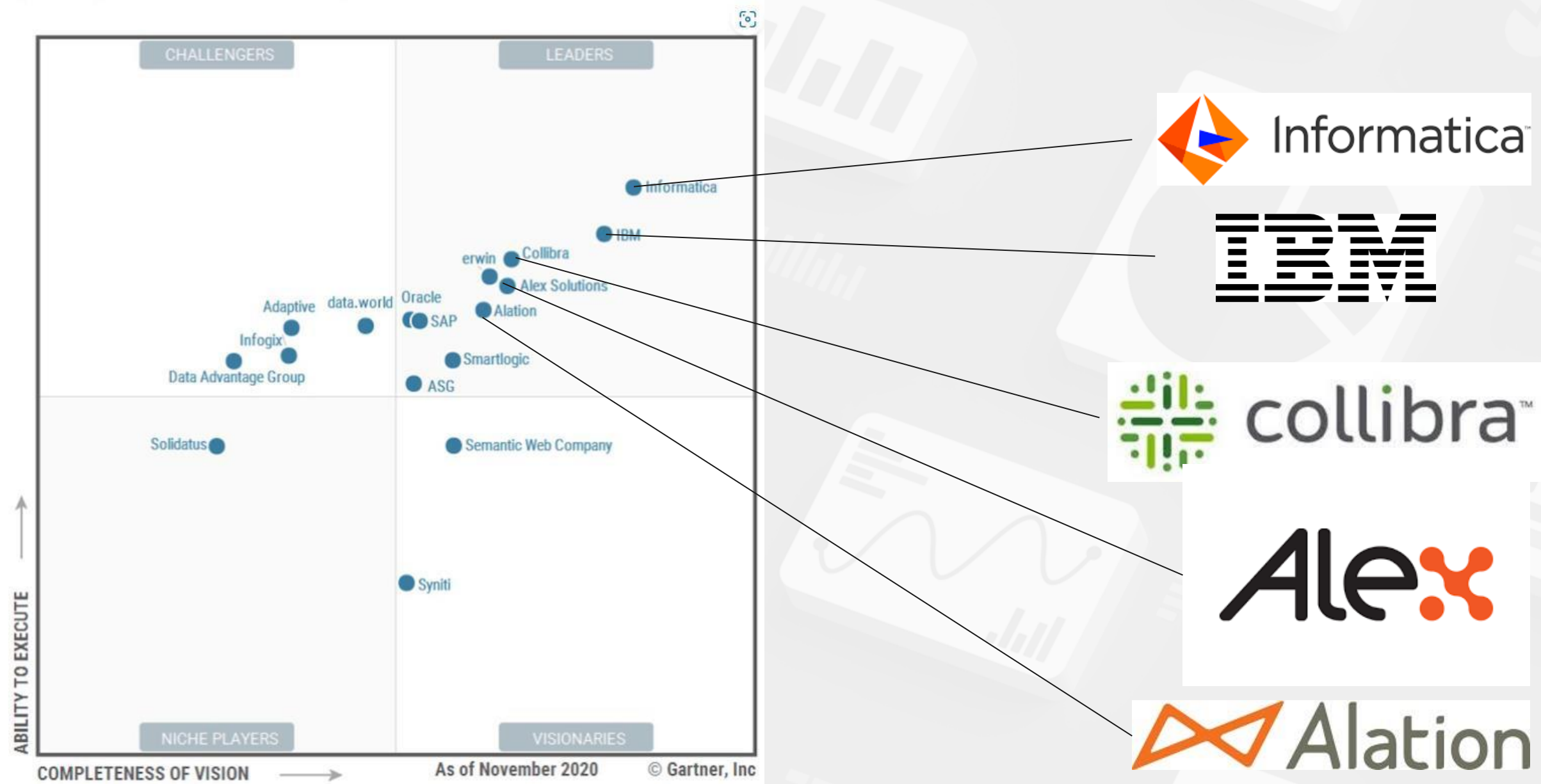
## SAP Power Designer

Con funcionalidades como la creación de modelos de datos, enlaces y sincronización y la gestión de metadatos, puede capturar automáticamente capas y requisitos de arquitectura, aprovechar un poderoso repositorio de metadatos y compartir descubrimientos. Características:

- Implementación on-premise
- Mapeo de la información e interfaz de drag and drop para encontrar necesidades de negocio
- Análisis de impacto de TI que descubre riesgos y obstáculos potenciales
- Visualizaciones intuitivas y paneles basados en la web para una mejor colaboración

<https://www.sap.com/spain/products/powerdesigner-data-modeling-tools.html>

Figure 1. Magic Quadrant for Metadata Management Solutions



Source: Gartner (November 2020)

# Los metadatos son tan esenciales para la gestión de datos no estructurados como lo son para la gestión de datos estructurados



A. Verdadero

B. Falso

# ¡Gracias!



## GOLD SPONSORS



## SILVER SPONSORS



## Entidades académicas

