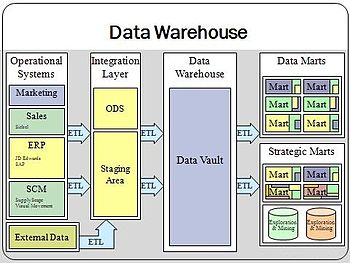
**Data Warehouse (Almacen de Datos)**

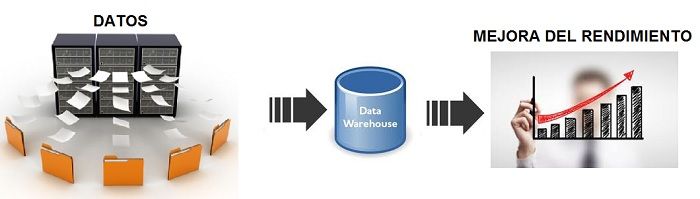
¿Qué es?

Una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza. Se usa para realizar informes (reports) y análisis de datos​ y se considera un componente fundamental de la inteligencia empresarial.​ Se trata, sobre todo, de un expediente completo de una organización, más allá de la información transaccional y operacional, almacenado en una base de datos diseñada para favorecer el análisis y la divulgación eficiente de datos (especialmente OLAP, procesamiento analítico en línea).

El almacenamiento de los datos no debe usarse con datos de uso actual. Los almacenes de datos contienen a menudo grandes cantidades de información que se subdividen a veces en unidades lógicas más pequeñas dependiendo del subsistema de la entidad del que procedan o para el que sea necesario.

¿Cómo se usa?

En un almacén de datos lo que se quiere es contener datos que son necesarios o útiles para una organización, es decir, que se utiliza como un repositorio de datos para posteriormente transformarlos en información útil para el usuario. Un almacén de datos debe entregar la información correcta a la gente indicada en el momento óptimo y en el formato adecuado. El almacén de datos da respuesta a las necesidades de usuarios expertos, utilizando Sistemas de Soporte a Decisiones (DSS), Sistemas de información ejecutiva (EIS) o herramientas para hacer consultas o informes. Los usuarios finales pueden hacer fácilmente consultas sobre sus almacenes de datos sin tocar o afectar la operación del sistema.



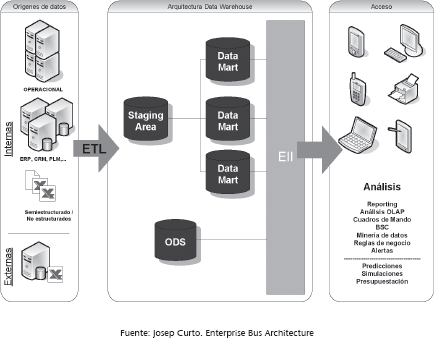
En el funcionamiento de un almacén de datos son muy importantes las siguientes ideas:

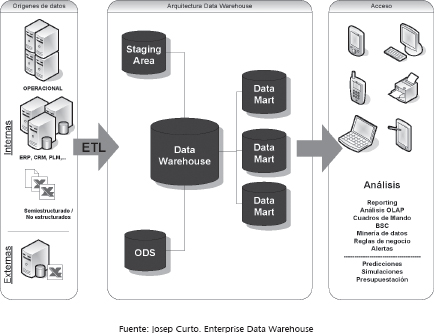
1. Integración de los datos provenientes de bases de datos distribuidas por las diferentes unidades de la organización y que con frecuencia tendrán diferentes estructuras (fuentes heterogéneas). Se debe facilitar una descripción global y un análisis comprensivo de toda la organización en el almacén de datos.
2. Separación de los datos usados en operaciones diarias de los datos usados en el almacén de datos para los propósitos de divulgación, de ayuda en la toma de decisiones, para el análisis y para operaciones de control. Ambos tipos de datos no deben coincidir en la misma base de datos, ya que obedecen a objetivos muy distintos y podrían entorpecerse entre sí.

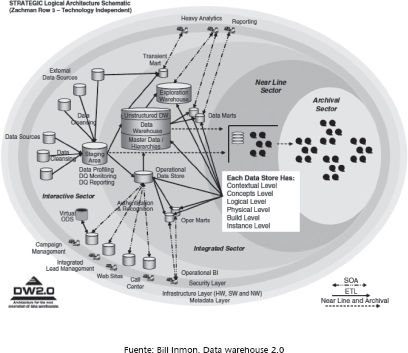
Periódicamente, se importan datos al almacén de datos de los distintos sistemas de planeamiento de recursos de la entidad (ERP) y de otros sistemas de software relacionados con el negocio para la transformación posterior. Es práctica común normalizar los datos antes de combinarlos en el almacén de datos mediante herramientas de extracción, transformación y carga (ETL). Estas herramientas leen los datos primarios (a menudo bases de datos OLTP de un negocio), realizan el proceso de transformación al almacén de datos (filtración, adaptación, cambios de formato, etc.) y escriben en el almacén.

¿Cómo se implementaría?

Existen principalmente tres enfoques en la arquitectura corporativa de un *data warehouse:*

• *Enterprise Bus Architecture* (o *Data Warehouse Virtual* / Federado): También conocido como MD *(Multidimensional Architecture),* consiste en una arquitectura basada en data marts independientes federados que pueden hacer uso de una staging area en caso de ser necesario. Federados quiere decir que se hace uso de una herramienta EII (*Enterprise Information Integration*) para realizar las consultas como si se tratara de un único data warehouse. Puede existir también un ODS, en caso de que se considere necesario.

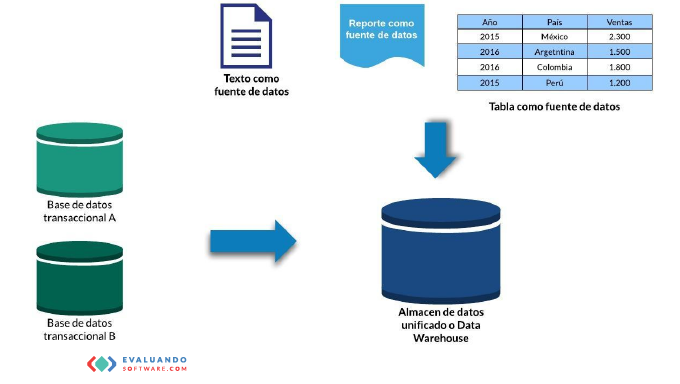
• *Corporate Information Factory* (o *Enterprise Data Warehouse):* Consiste en una arquitectura en la que existe un data warehouse corporativo y distintos data mart (o incluso cubos OLAP) dependientes del mismo. El acceso a datos se realiza a partir de los data marts o del ODS en caso de existir, pero nunca se consulta el data warehouse directamente. Puede existir, en el caso de ser necesaria, una staging area.



• *Enterprise Data Warehouse 2.0:* consiste en la revisión de la metodología de Bill Inmon para incluir toda la experiencia de los últimos veinte años. El punto diferencial es que se separa la información según su antigüedad y se clasifica según su uso. Se caracteriza por incluir tanto información estructurada como no estructurada y por focalizarse en responder a todas las necesidades actuales de negocio. Es decir, distingue diferentes entornos en función de lo actual que es la información y como se usa.

**Metadatos**

Uno de los componentes más importantes de la arquitectura de un almacén de datos son los metadatos. Se define comúnmente como "datos acerca de los datos", en el sentido de que se trata de datos que describen cuál es la estructura de los datos que se van a almacenar y cómo se relacionan.

El metadato documenta, entre otras cosas, qué tablas existen en una base de datos, qué columnas posee cada una de las tablas y qué tipo de datos se pueden almacenar. Los datos son de interés para el usuario final, el metadato es de interés para los programas que tienen que manejar estos datos. Sin embargo, el rol que cumple el metadato en un entorno de almacén de datos es muy diferente al rol que cumple en los ambientes operacionales. En el ámbito de los data warehouse el metadato juega un papel fundamental, su función consiste en recoger todas las definiciones de la organización y el concepto de los datos en el almacén de datos, debe contener toda la información concerniente a:

* Tablas
* Columnas de tablas
* Relaciones entre tablas
* Jerarquías y Dimensiones de datos
* Entidades y Relaciones

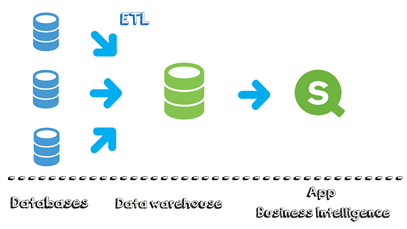
**Procesos de ETL**

En el contexto del data warehousing existen diferentes técnicas y tecnologías para la integración de datos. Se entiende por integración de datos al conjunto de aplicaciones, productos, técnicas y tecnologías, que permiten una visión única consistente de nuestros datos de negocio.

En el contexto de la inteligencia de negocio, las herramientas de ETL han sido la opción usual para alimentar el data warehouse. La funcionalidad básica de estas herramientas está compuesta por:

* Extracción. Acción de obtener la información deseada a partir de los datos almacenados en fuentes externas.
* Transformación. Cualquier operación realizada sobre los datos para que puedan ser cargados en el data warehouse o se puedan migrar de éste a otra base de datos.
* Carga. Consiste en almacenar los datos en la base de datos final, por ejemplo el almacén de datos objetivo normal.

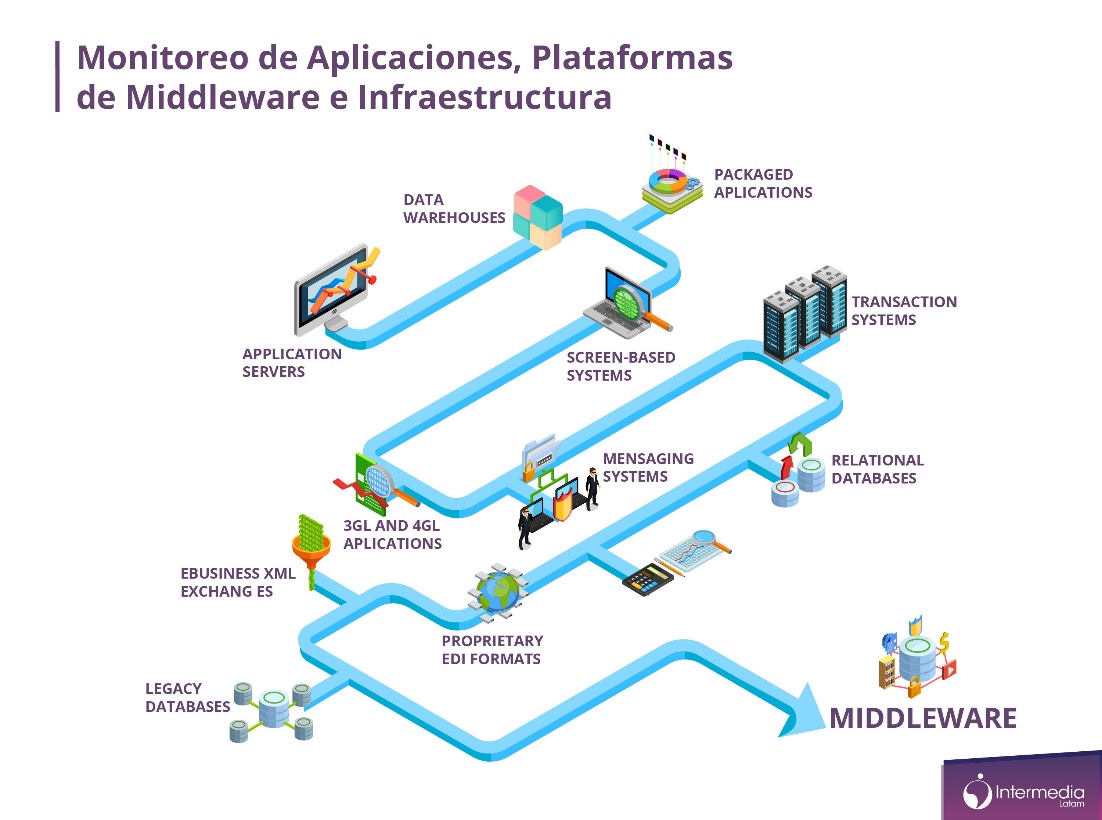
En los últimos años, estas herramientas han evolucionado incluyendo más funcionalidades propias de una herramienta de integración de datos. Podemos destacar:

* Servicios de acceso/entrega de datos (vía adaptadores/conectores)
* Gestión de servicios
* Data profiling
* Data Quality
* Procesos Operacionales
* Servicios de transformación: CDC, SCD, validación, agregación
* Servicios de acceso a tiempo real.
* Extract, Transform and Load (ETL)
* Enterprise Information Integration (EII)
* Enterprise Application Integration (EAI)
* Capa de transporte de datos
* Gestión de metadatos

Esta evolución es consecuencia de diversos motivos. Por un lado, está la propia evolución de las necesidades de negocio y por otro, el incremento en riqueza de los tipos de datos disponibles en la organización:

* Estructurados: contenidos en bases de datos.
* Semiestructurados: en formatos legibles para máquinas si bien no están completamente estructurados: HTML tabulado, Excel, CSV... que pueden obtenerse mediante técnicas estándar de extracción de datos.
* No estructurados: en formatos legibles para humanos, pero no para máquinas: Word, HTML no tabulado, PDF. que pueden obtenerse mediante técnicas avanzadas como text mining u otras.

**Middleware**

Término genérico que se utiliza para referirse a todo tipo de software de conectividad que ofrece servicios u operaciones que hacen posible el funcionamiento de aplicaciones distribuidas sobre plataformas heterogéneas. Estos servicios funcionan como una capa de abstracción de software distribuida, que se sitúa entre las capas de aplicaciones y las capas inferiores (sistema operativo y red). El middleware puede verse como una capa API, que sirve como base a los programadores para que puedan desarrollar aplicaciones que trabajen en diferentes entornos sin preocuparse de los protocolos de red y comunicaciones en que se ejecutarán. De esta manera se ofrece una mejor relación costo/rendimiento que pasa por el desarrollo de aplicaciones más complejas, en menos tiempo.

La función del middleware en el contexto de los data warehouse es la de asegurar la conectividad entre todos los componentes de la arquitectura de un almacén de datos.

¿Como funciona en mi empresa?

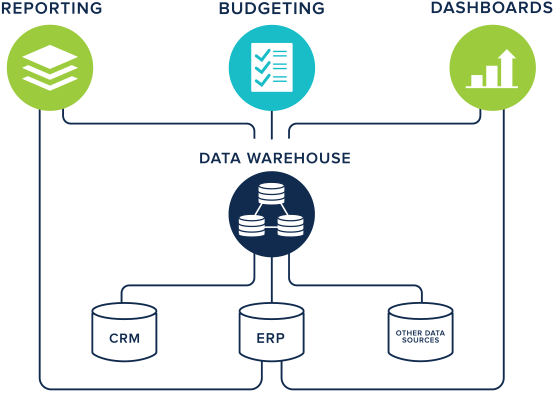
La creación del Data Warehouse representa la mayoría de las veces el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de Business Intelligence y así aportar las mejores respuestas a los problemas de la organización.

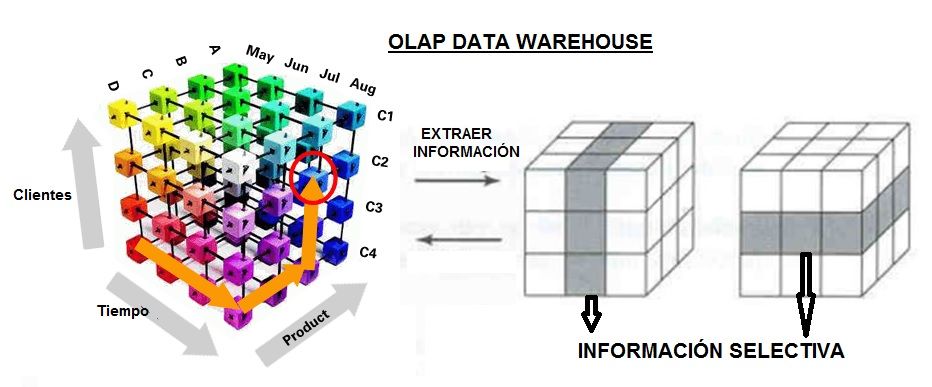
Aportaciones del proceso:

1. Nos proporciona información clave para la toma de decisiones en la empresa.
2. Mejora la calidad de las decisiones tomadas.
3. Estos datos son útiles a mediano y largo plazo.
4. Si las fuentes de los datos y los objetivos están claros, los sistemas son sencillos de instalar.
5. Es útil también para el almacenamiento de análisis y consultas históricas.
6. Permite mayor flexibilidad y rapidez en el acceso a la información.
7. Brinda una comunicación fiable entre todos los departamentos de la empresa.

También hay que destacar que, para comprender íntegramente el concepto, este proceso se construye a partir del ETL (Extracción, transformación y carga) y los sistemas operacionales de una compañía.

P ara que el Data Warehouse sea exitoso se necesita una limpieza continua, transformación e integración de los datos. Además, requiere sistemas, aplicaciones y almacenamiento específico. Es un trabajo constante que garantiza el éxito de los datos en el diagnóstico y las soluciones de inteligencia de negocios que serán implementadas en la compañía.





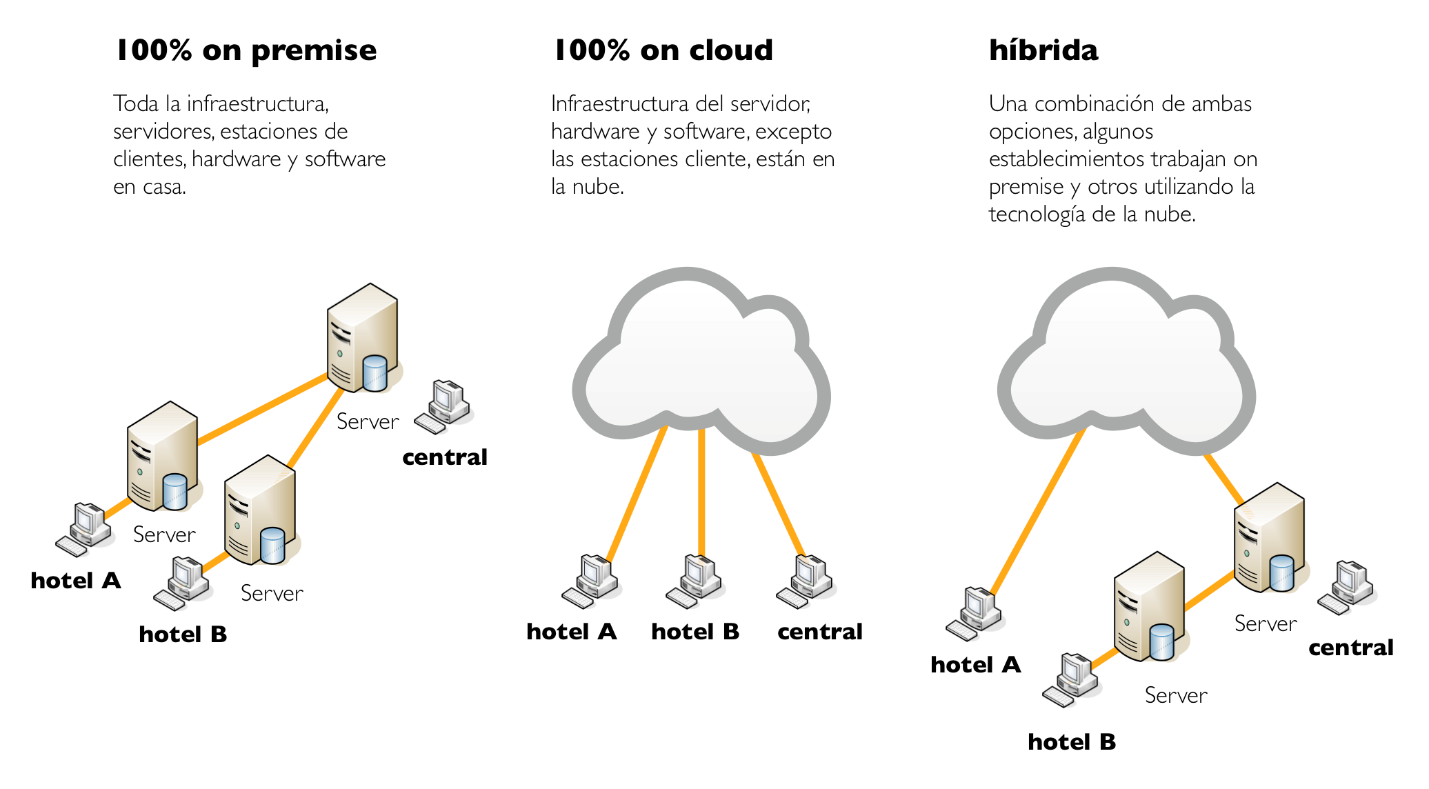
Ejemplos de Data warehouse:

MS SQL SERVER Business Inteligence es la solución para construir data warehouse del manejador de base de datos SQL SERVER.

EXCEL Business Inteligence: La hoja de cálculo EXCEL de Microsoft permite la construcción de cubos y almacenes de datos para realizar análisis de información por medio de escenarios.

Oracle Data Warehousing: Es la solución de Oracle para la creación de data warehouse.

**Sistemas On Premise**

¿Qué es?

Es un sistema “in house” alojada para el cliente, su instalación, integración, administración y mantenimiento es llevaba a cabo por especialistas IT. Este sistema es gestionado por los propios empleados.

Es habitual ver este tipo de soluciones On Premise en empresas grandes debido a su mayor personalización y para evitar sobrecargar el servidor por carga de información.

Ofrecen una gestión controlada con flujos de trabajo predefinidos y una configuración y formación personalizada al modelo de la empresa. Es decir cada sistema está diseñado a las necesidades de la empresa

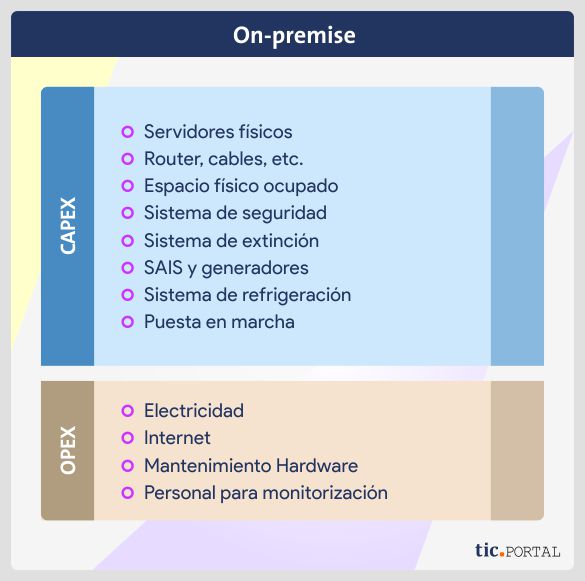
¿Cómo se usa?

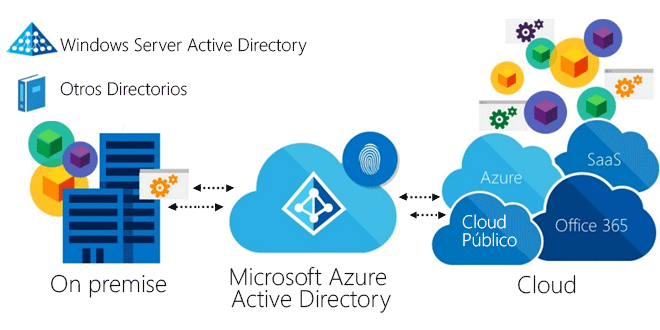
Esta instalación se lleva a cabo dentro del servidor y la infraestructura (TIC) de la empresa. Es el modelo tradicional de aplicaciones empresariales.

Con el modelo on-premise, la empresa es la responsable de la seguridad, disponibilidad y gestión del software. Por lo que la empresa debe tener un departamento de sistemas que dedique parte de sus recursos a la gestión de la infraestructura in situ. Sin embargo, el proveedor también suele proporcionar servicios de integración y soporte post-venta. Las soluciones TIC que pueden encontrarse en este tipo de implementación son aquellas que ya llevan un tiempo en el mercado como gestión documental, sistemas ERP o gestión de relaciones con el cliente (CRM), las más nuevas suelen ofrecerse solamente en la nube, que es otro tipo de instalación.

La instalación on-premise ofrece ventajas como un mayor control, pero la inversión inicial es más arriesgada y muchas soluciones on-premise no soportan dispositivos móviles ni wearables. En la siguiente tabla se comparan los pros y contras de este tipo de instalación:

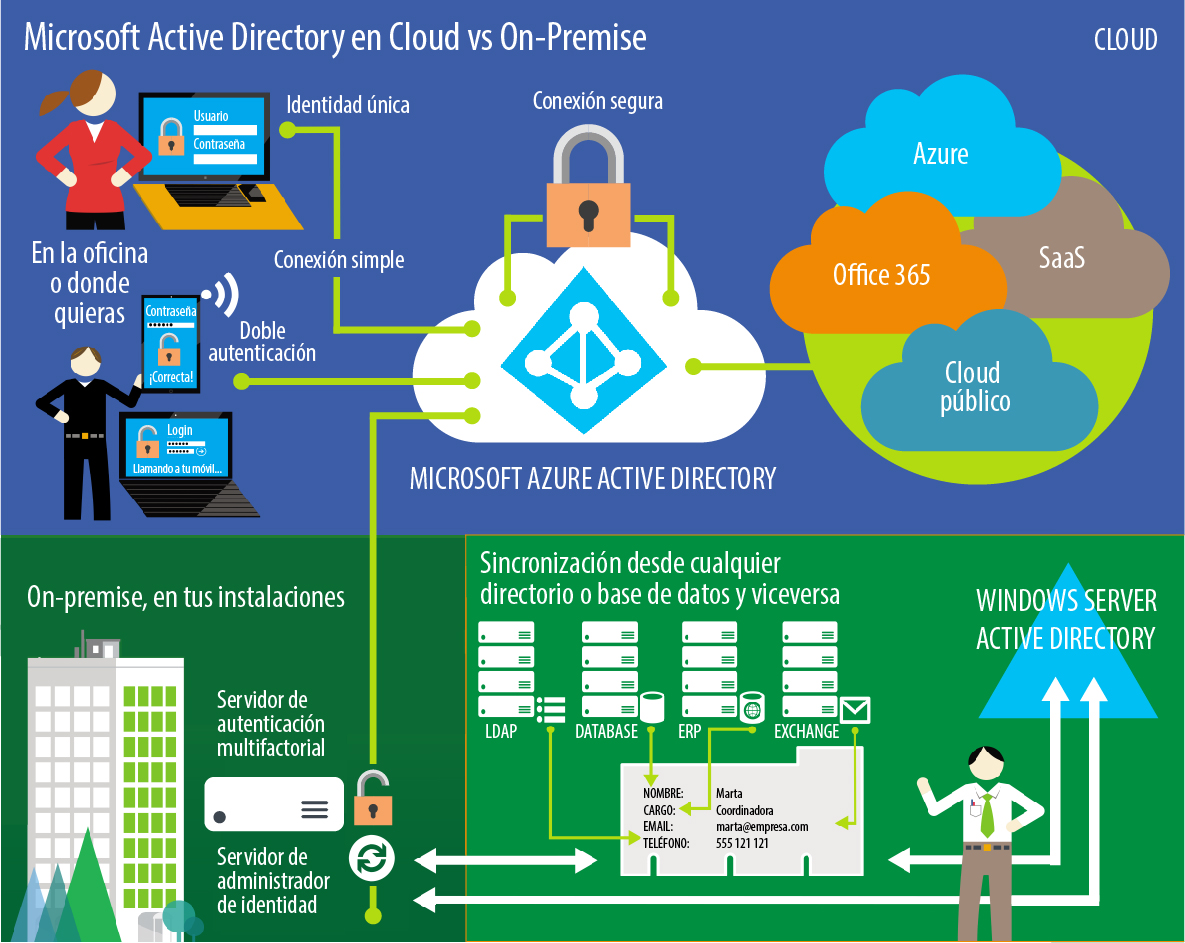
¿Cómo se implementaría?

Un proyecto de implementación de ERP se divide en varias fases. ﻿Cada una de estas fases se refiere a competencias específicas:

* Estudio previo: Formalización de las normas de gestión, revisión de estas normas, formalización de los requerimientos funcionales. Esta fase requiere competencias funcionales y departamentales. A menudo la llevan a cabo consultores funcionales en estrecha colaboración con los empleados del departamento en cuestión.
* Desarrollos/configuración, redacción del plan de parametrizaciones.
* Preparación y realización de pruebas (pruebas de integración formales, pruebas de compatibilidad ascendente).
* Capacitación de los usuarios y apoyo hacia la transición.
* Transferencia desde el sistema antiguo al nuevo, control de la recuperación correcta de datos (ejemplo: (asientos contables iniciales).

El tipo de implementación de un ERP dependerá de los recursos con los que cuente tu empresa. Normalmente se necesita al menos un recurso IT dedicado en la empresa para una efectiva implantación ERP, dado las integraciones con otros softaware necesarias. Las soluciones on premise suelen ofrecer más nivel de personalización que las soluciones online, aunque esto está cambiando últimamente.

Por otra parte, el licenciamiento está basado en la compra de la solución y, dependiendo del proveedor, en el pago anual de un mantenimiento y soporte. Esta opción también ofrece más flexibilidad en la configuración del ERP en conjunto con el proveedor del ERP.

¿Como funcionaría?