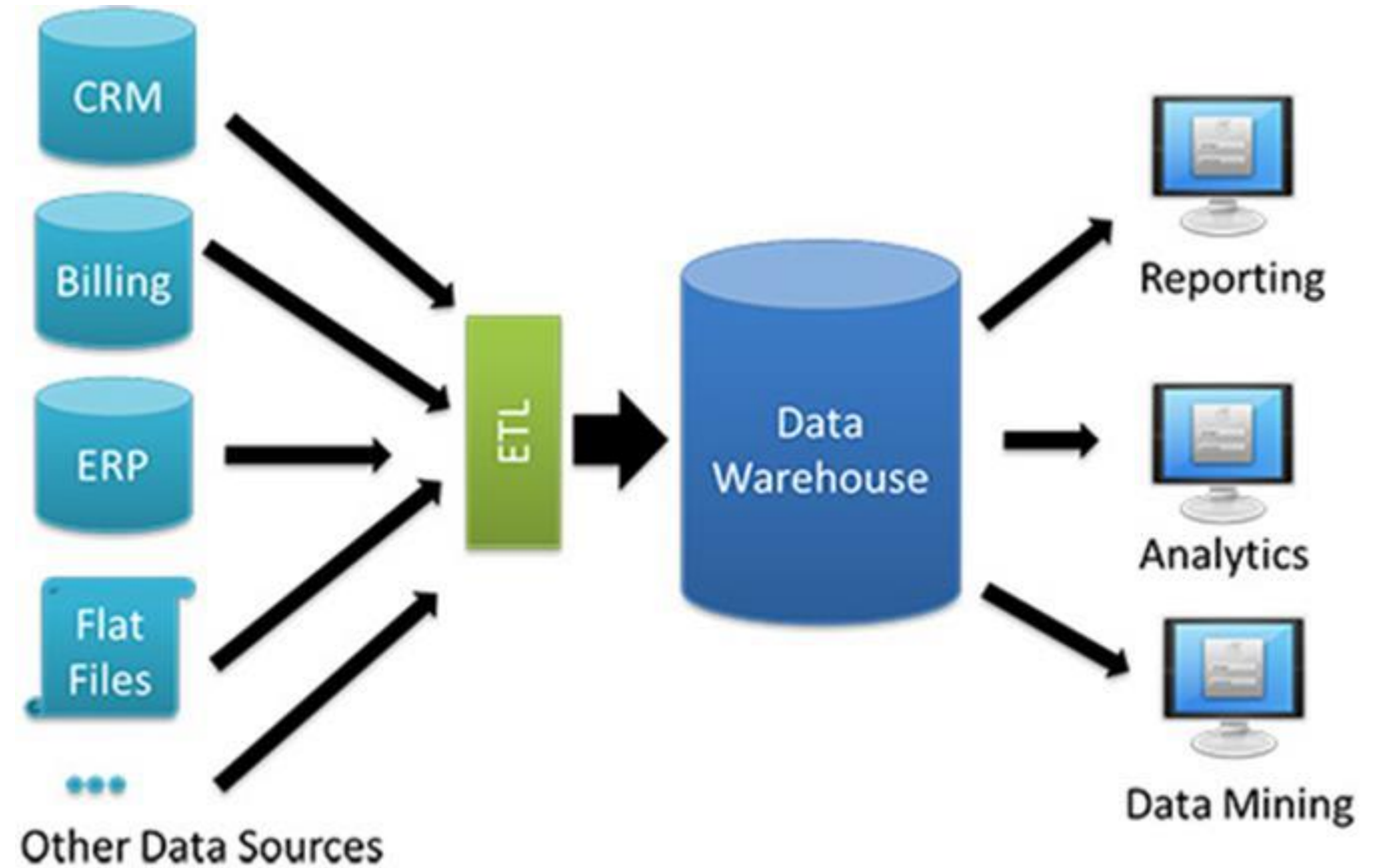


INTRODUCCION AL DATA WAREHOUSE

Un data warehouse es un repositorio unificado para todos los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso.

Normalmente, un data warehouse se aloja en un servidor corporativo o en la nube. Los datos de diferentes aplicaciones de procesamiento de transacciones Online (OLTP) y otras fuentes se extraen selectivamente para su uso por aplicaciones analíticas y de consultas por usuarios.

Data Warehouse es una arquitectura de almacenamiento de datos que permite a los ejecutivos de negocios organizar, comprender y utilizar sus datos para tomar decisiones estratégicas.

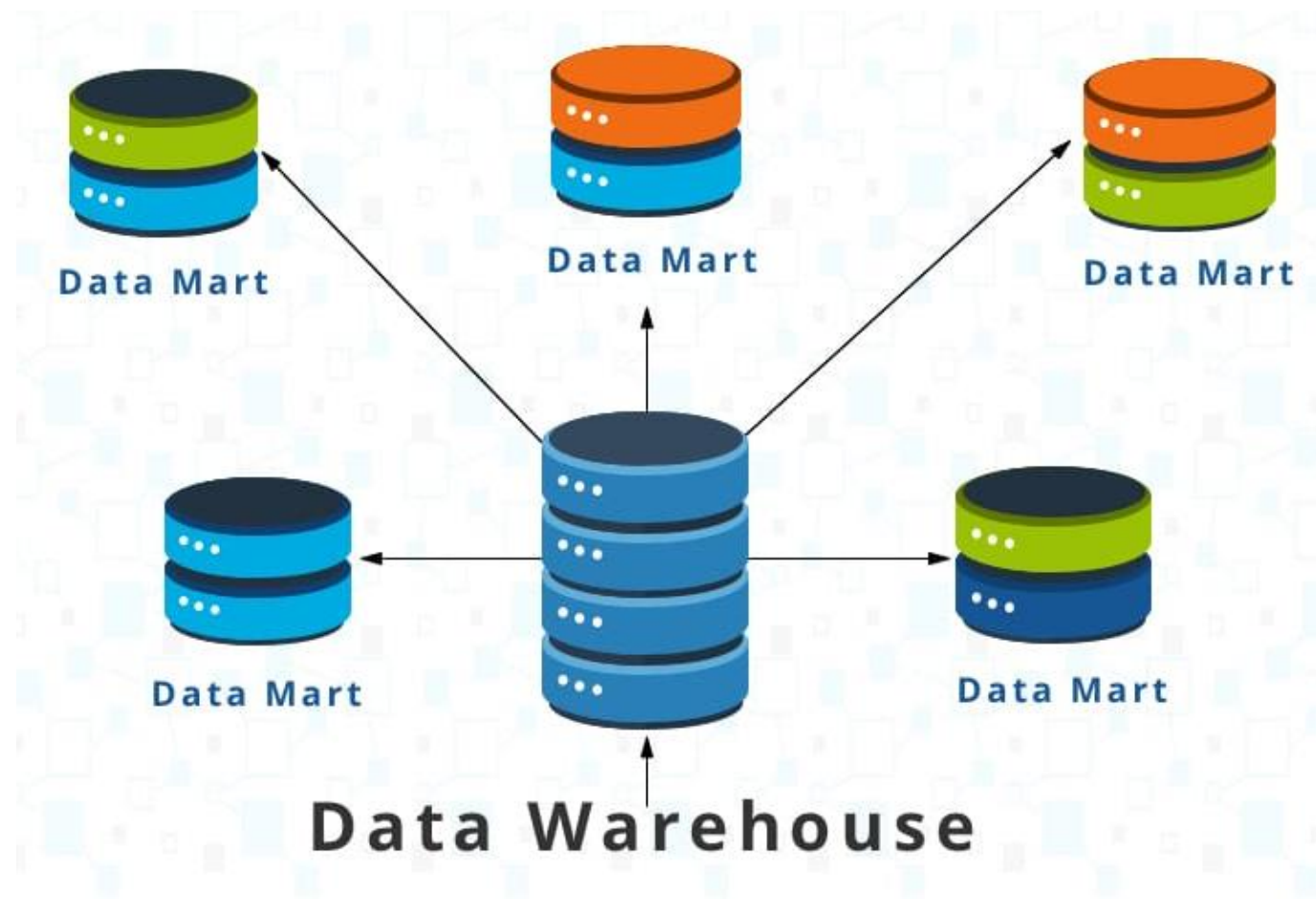


DATAMART

Un Datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un datamart puede ser alimentado desde los datos de un datawarehouse, o integrar por si mismo un compendio de distintas fuentes de información.

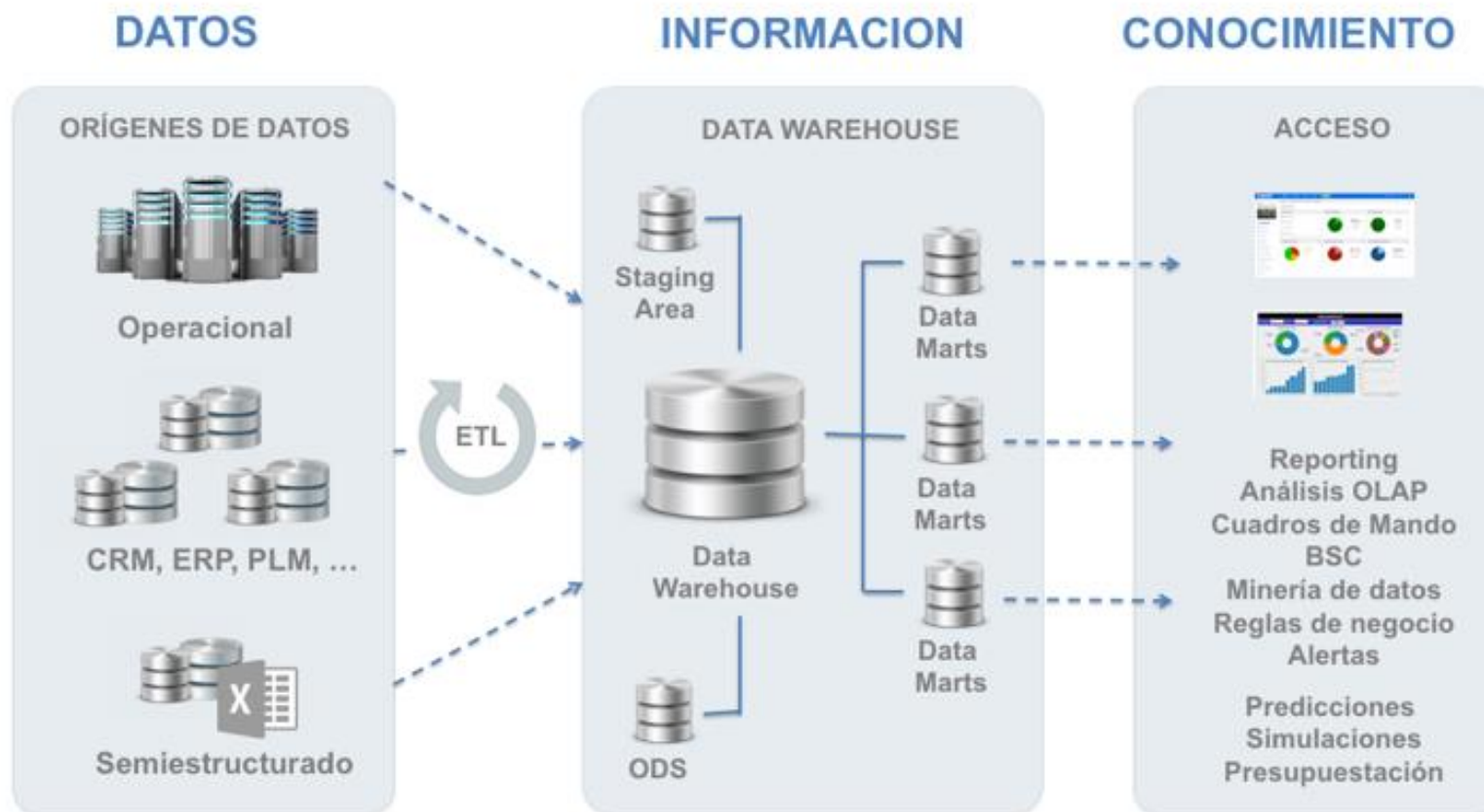
Los datamarts que están dotados con estas estructuras óptimas de análisis presentan las siguientes ventajas:

- › Poco volumen de datos
- › Mayor rapidez de consulta
- › Consultas SQL y/o MDX sencillas
- › Validación directa de la información
- › Facilidad para la historización de los datos



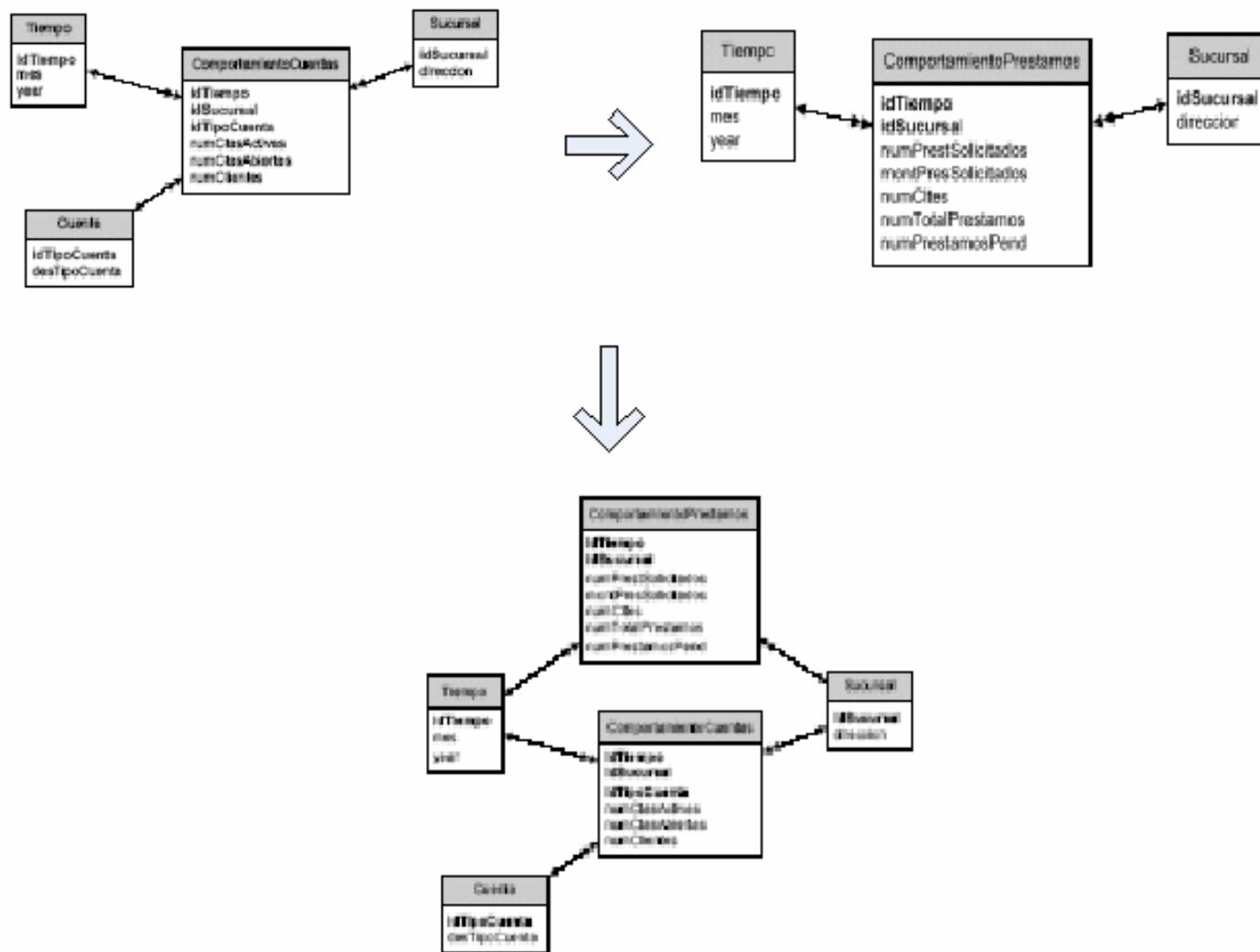
CARACTERÍSTICAS DE UN DATA WAREHOUSE

- › Integrado: los datos almacenados en el Data Warehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.



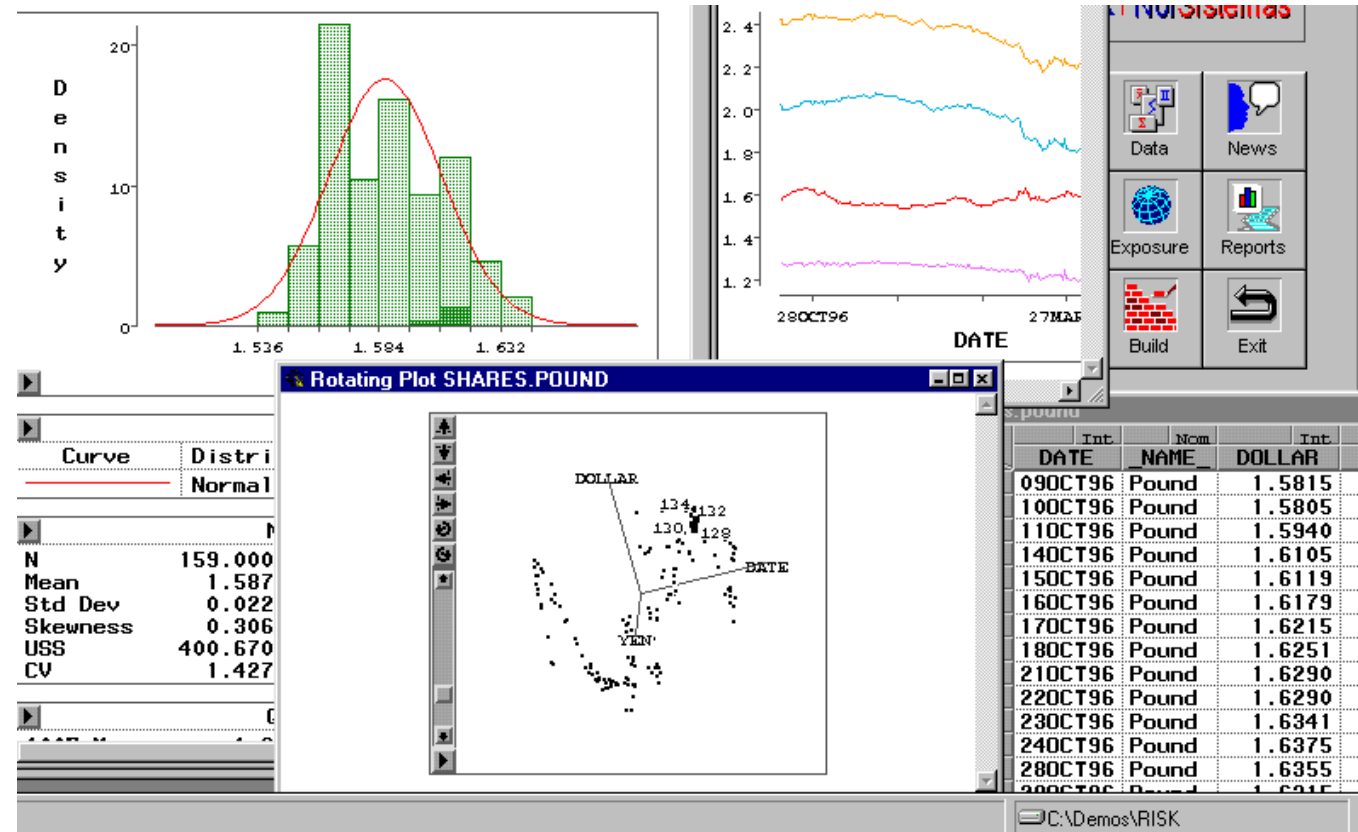
CARACTERISTICAS DE UN DATA WAREHOUSE

- › Temático: sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del Data Warehouse. De esta forma, las peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.



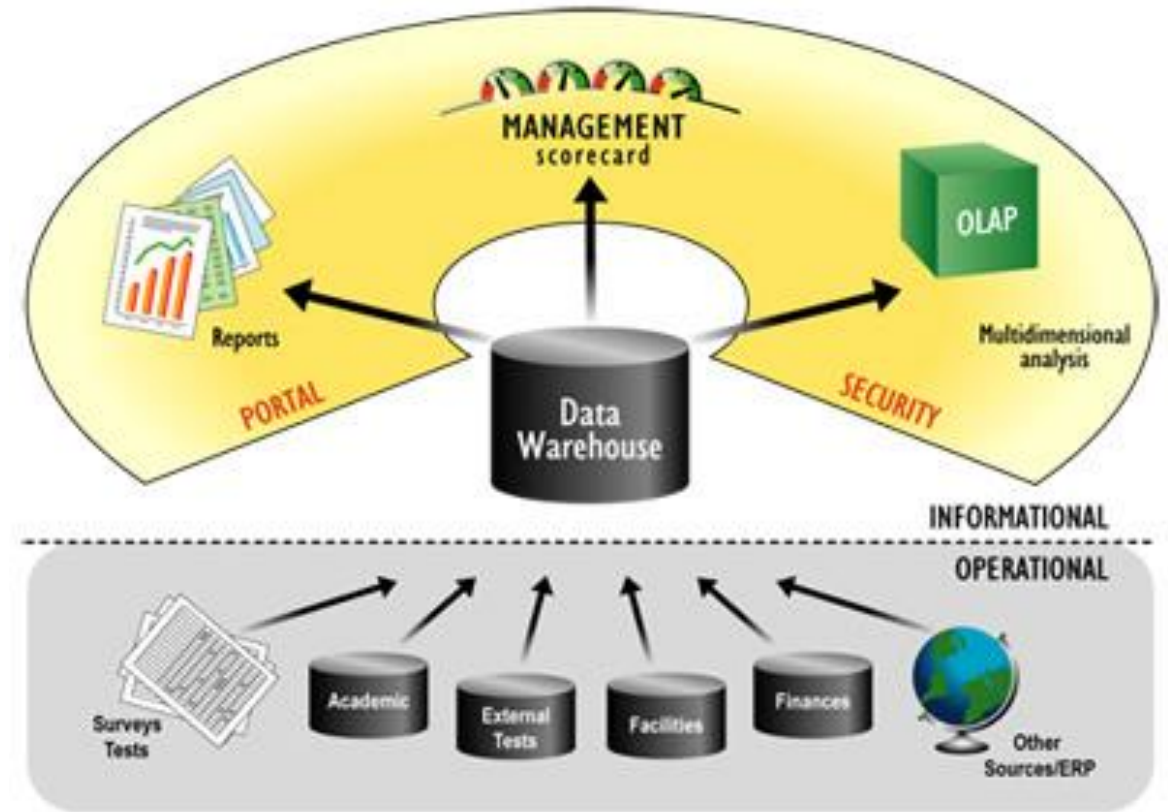
CARACTERISTICAS DE UN DATA WAREHOUSE

- › Histórico: el tiempo es parte implícita de la información contenida en un Data Warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el Data Warehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el Data Warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

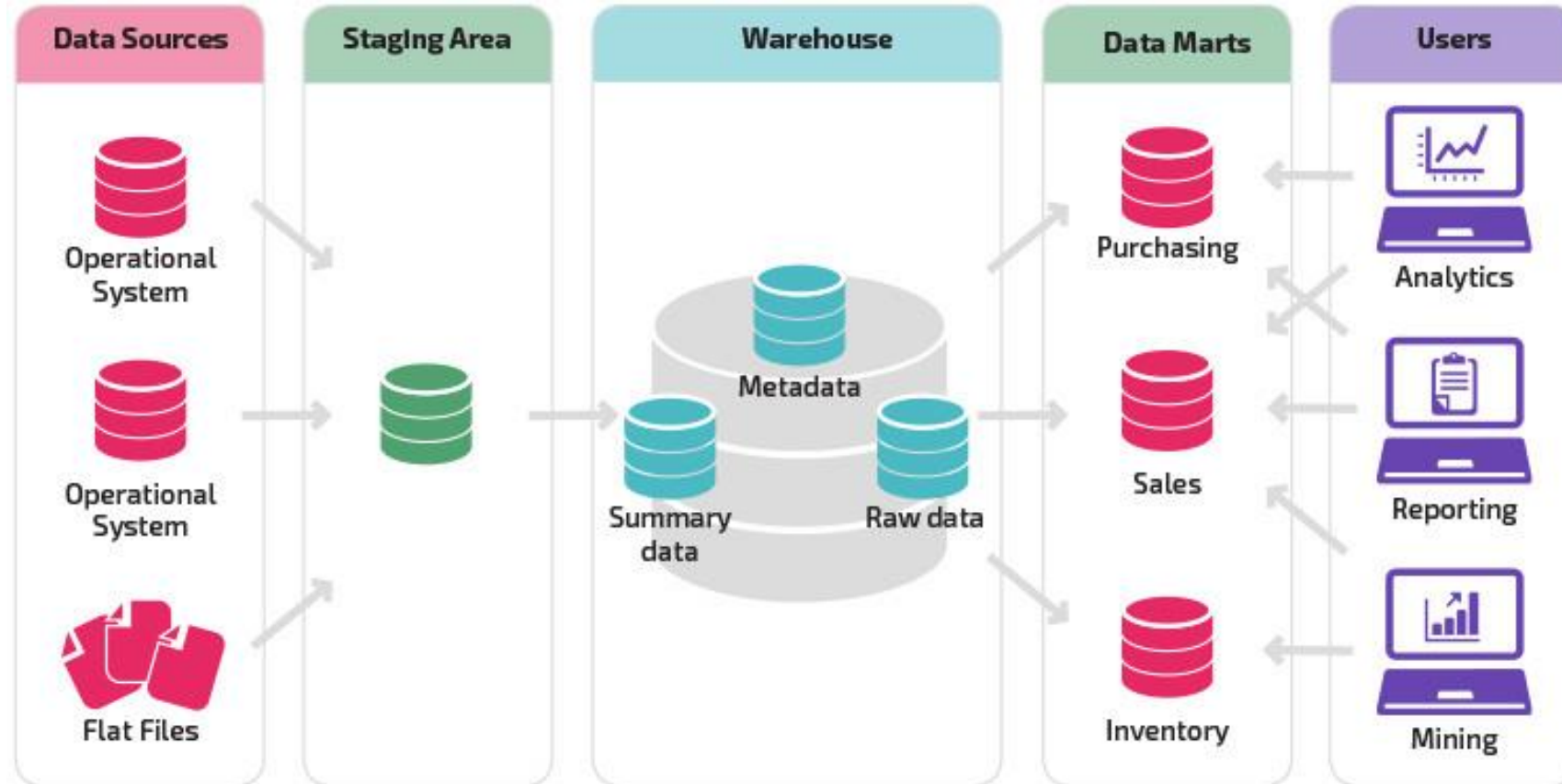


CARACTERISTICAS DE UN DATA WAREHOUSE

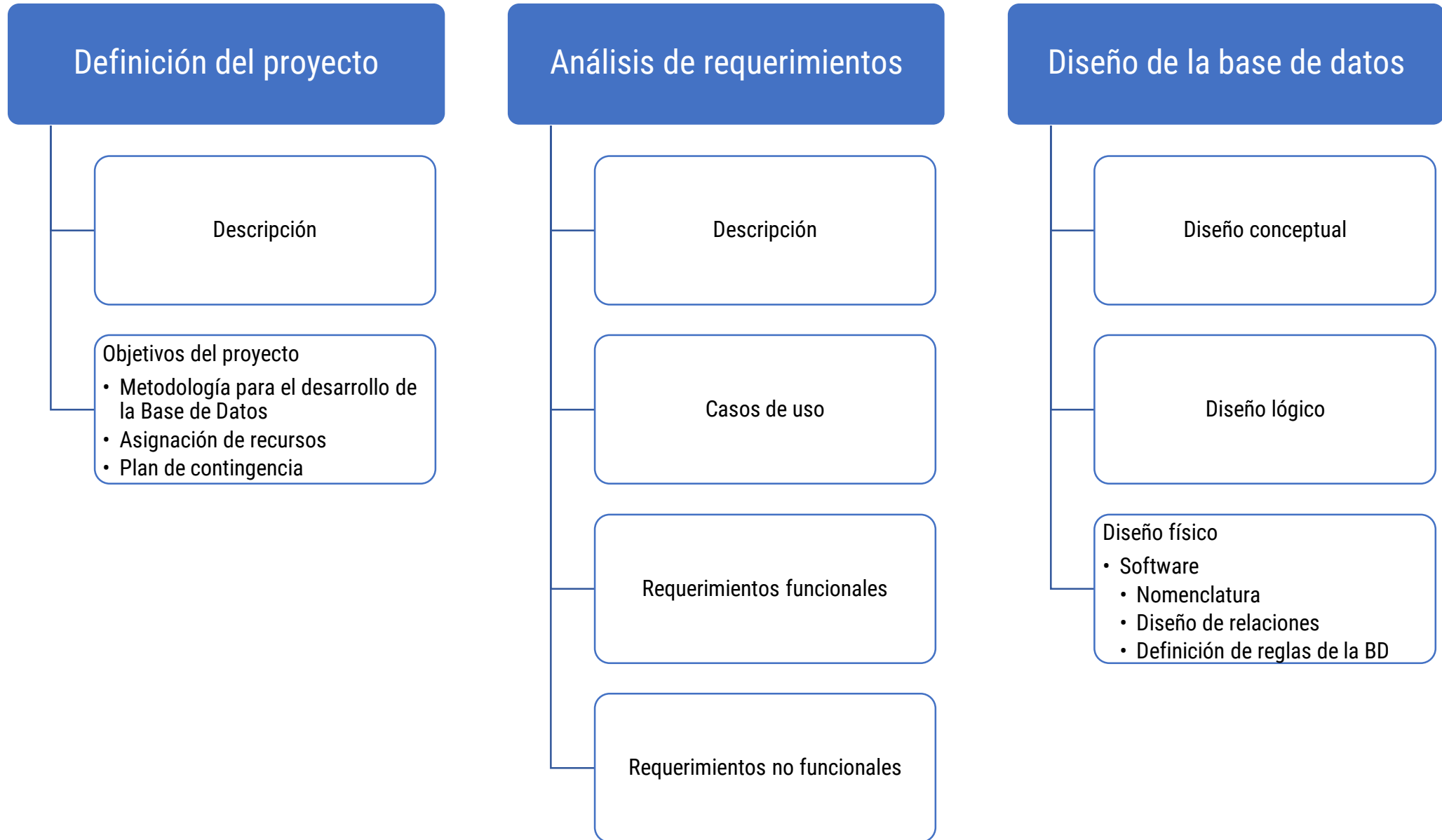
- › No volátil: el almacén de información de un Data Warehouse existe para ser leído, y no modificado. La información es por tanto permanente, significando la actualización del Data Warehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.



ARQUITECTURA DE UN DATA WAREHOUSE



PLANIFICACION DE INFRAESTRUCTURA DE UN DATA WAREHOUSE



PLANIFICACION DE INFRAESTRUCTURA DE UN DATA WAREHOUSE

Comprensión del negocio Se trata de entender claramente los requerimientos y objetivos del proyecto siempre desde una visión de negocio. Esta fase se subdivide a su vez en las siguientes categorías:

- › Definición de los objetivos de negocio y criterios de éxito del negocio (inicial, objetivos de negocio y criterios de éxito del negocio).
- › Evaluación de la situación (inventario de recursos, requisitos supuestos y requerimientos, riesgos y contingencias, terminología y costes y beneficios)
- › Definición de los objetivos del Data Warehouse (objetivos y criterios de éxito).
- › Realización del plan del proyecto (plan del proyecto y valoración inicial de herramientas y técnicas).

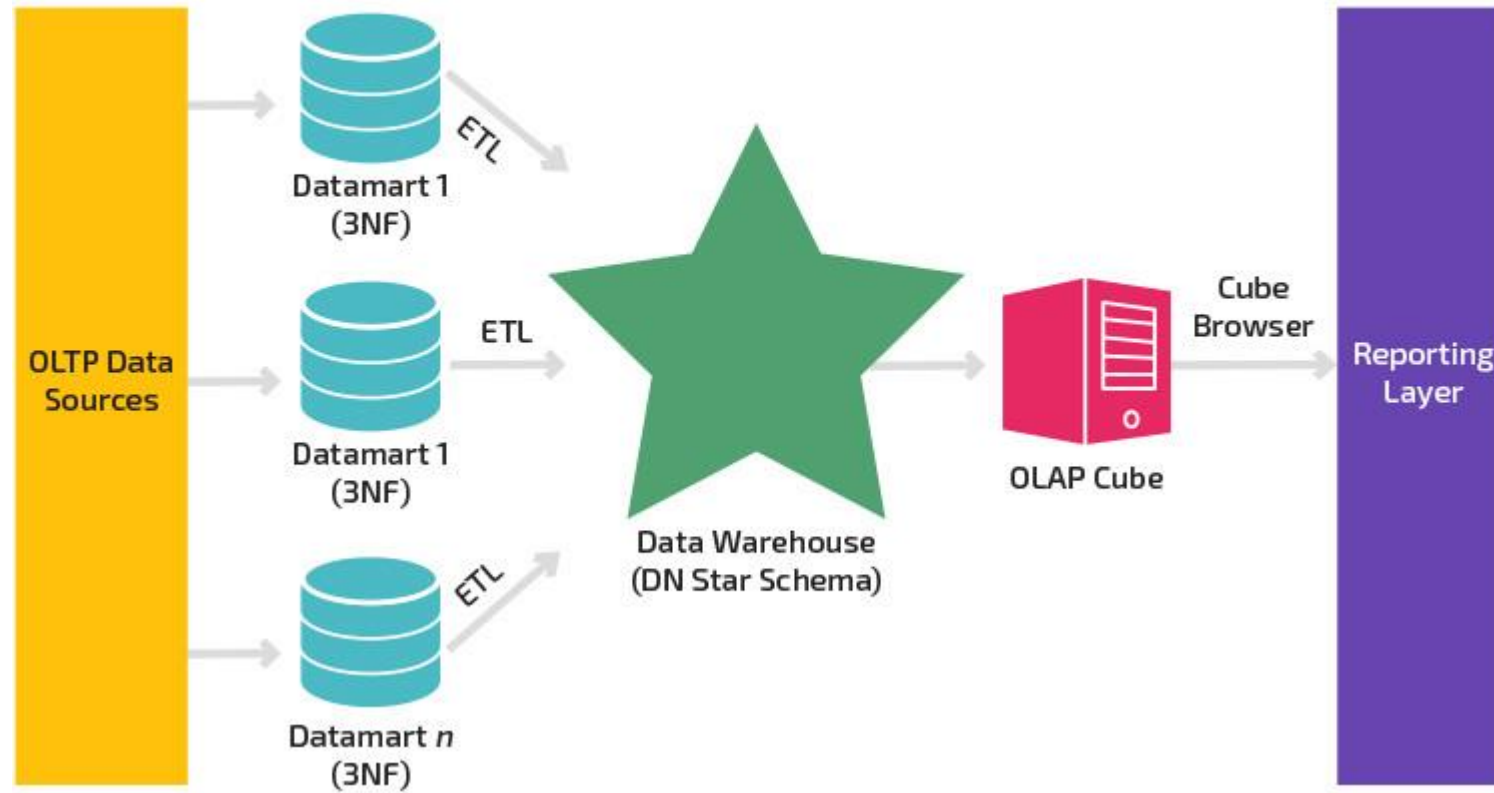
METODOLOGIA DE KIMBALL PARA EL DATA WAREHOUSE

La metodología se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle). Este ciclo de vida del proyecto de DW, está basado en cuatro principios básicos:

- › Centrarse en el negocio: Hay que concentrarse en la identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio, agudizando el análisis del mismo y la competencia consultiva de los implementadores.
- › Construir una infraestructura de información adecuada: Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.
- › Realizar entregas en incrementos significativos: crear el almacén de datos (DW) en incrementos entregables en plazos de 6 a 12 meses. Hay que usar el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos. En esto la metodología se parece a las metodologías ágiles de construcción de software.
- › Ofrecer la solución completa: proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible. También se deberá entregar herramientas de consulta ad hoc, aplicaciones para informes y análisis avanzado, capacitación, soporte, sitio web y documentación.

METODOLOGIA DE KIMBALL PARA EL DATA WAREHOUSE

Kimball Model



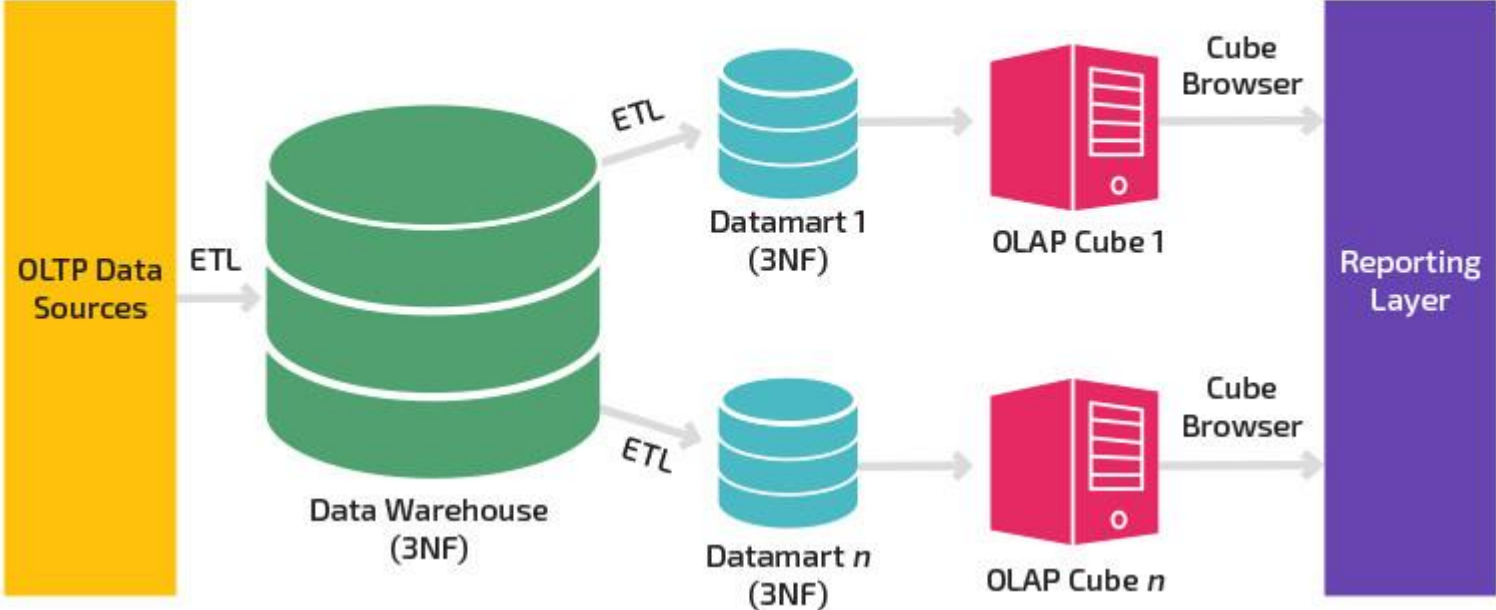
METODOLOGIA DE INMON PARA EL DATA WAREHOUSE

En este modelo, la premisa es que la información se almacene al máximo nivel de detalle (garantizando la futura exploración de los datos), permaneciendo invariable y no volátil, de manera que los cambios que sufran los datos a lo largo del tiempo queden registrados sin que puedan modificarse o eliminarse.

En este modelo, la premisa es que la información se almacene al máximo nivel de detalle (garantizando la futura exploración de los datos), permaneciendo invariable y no volátil, de manera que los cambios que sufran los datos a lo largo del tiempo queden registrados sin que puedan modificarse o eliminarse.

Estas son las claves fundamentales de la arquitectura defendida por Inmon, conocida como 'Corporate Information Factory (CIF)', donde el DataWarehouse centraliza todos los datos de la compañía para alimentar, a continuación, pequeños DataMarts temáticos, que serán los puntos de acceso para las herramientas de reporting. En este sentido, cada departamento tendrá su propio DataMart, abastecido con la información del DataWarehouse, listo para su análisis y explotación.

Inmon Model



DIFERENCIAS ENTRE KIMBALL E INMON

	Inmon	Kimball
Presupuesto	Coste inicial alto	Coste inicial bajo
Plazos	Requiere más tiempo de desarrollo	Tiempo de desarrollo inferior
Expertise	Equipo con especialización alta	Equipo con especialización media
Alcance	Toda la compañía	Departamentos individuales
Mantenimiento	Fácil mantenimiento	Mantenimiento más complejo