



Universidad
Nacional de
Cajamarca
"Norte de la Universidad Peruana"

Inteligencia de Negocios

03. Fundamentos de DataWarehouse.



Universidad Nacional de Cajamarca



www.unc.edu.pe/



Universidad Nacional de Cajamarca



Universidad
Nacional de
Cajamarca
"Norte de la Universidad Peruana"

INTRODUCCIÓN

03. Fundamentos de DataWarehouse





DATAWAREHOUSE

Las organizaciones a nivel global recopilan datos de manera continua para mejorar sus estrategias de inteligencia de negocios y la generación de informes.

Un Almacén de Datos, o Data Warehouse, permite a las empresas gestionar eficientemente esta información, asegurando que las operaciones diarias se lleven a cabo sin interrupciones.

Un Data Warehouse centraliza y almacena grandes cantidades de datos de diferentes fuentes, permitiendo a las organizaciones mejorar su rendimiento y tomar decisiones informadas basadas en análisis precisos y detallados

Para llevar a cabo BI, es esencial gestionar, depurar e integrar datos de diversas fuentes y formatos, almacenándolos en una base de datos central para su análisis y exploración. Este proceso se denomina Data Warehousing.



RECORDEMOS



Universidad
Nacional de
Cajamarca
"Norte de la Universidad Peruana"

- ¿Qué es la toma de decisiones?
- ¿Cuál es el ciclo de BI?
- ¿Qué es un KPI?
- ¿Porqué usar BI?



LOGRO ESPERADO

- Al finalizar la sesión, los estudiantes podrán comprender y explicar los fundamentos del Data Warehouse (DW), identificar las características clave del DW, y aplicar conceptos teóricos a un análisis práctico de una empresa local utilizando herramientas de inteligencia de negocios





Universidad
Nacional de
Cajamarca
"Norte de la Universidad Peruana"

DESARROLLO DEL TEMA

03. Fundamentos de DataWarehouse



Universidad Nacional de Cajamarca



www.unc.edu.pe/



Universidad Nacional de Cajamarca

¿Qué es DataWarehouse?

Un Data Warehouse es un sistema de bases de datos centralizado a nivel empresarial, diseñado específicamente para recolectar, almacenar e integrar datos de diversos sistemas operativos, poniéndolos a disposición para la generación de reportes y apoyo en la toma de decisiones empresariales.

Los datos se obtienen de una amplia variedad de fuentes relevantes para el negocio, tales como datos de clientes, datos de la competencia y datos de transacciones



Definiciones de Data Warehouse

- “Un Data Warehouse es un conjunto integrado de bases de datos, con orientación temática, que están diseñados para el apoyo a la Toma de Decisiones, y donde cada unidad de datos es relevante en algún momento del tiempo”

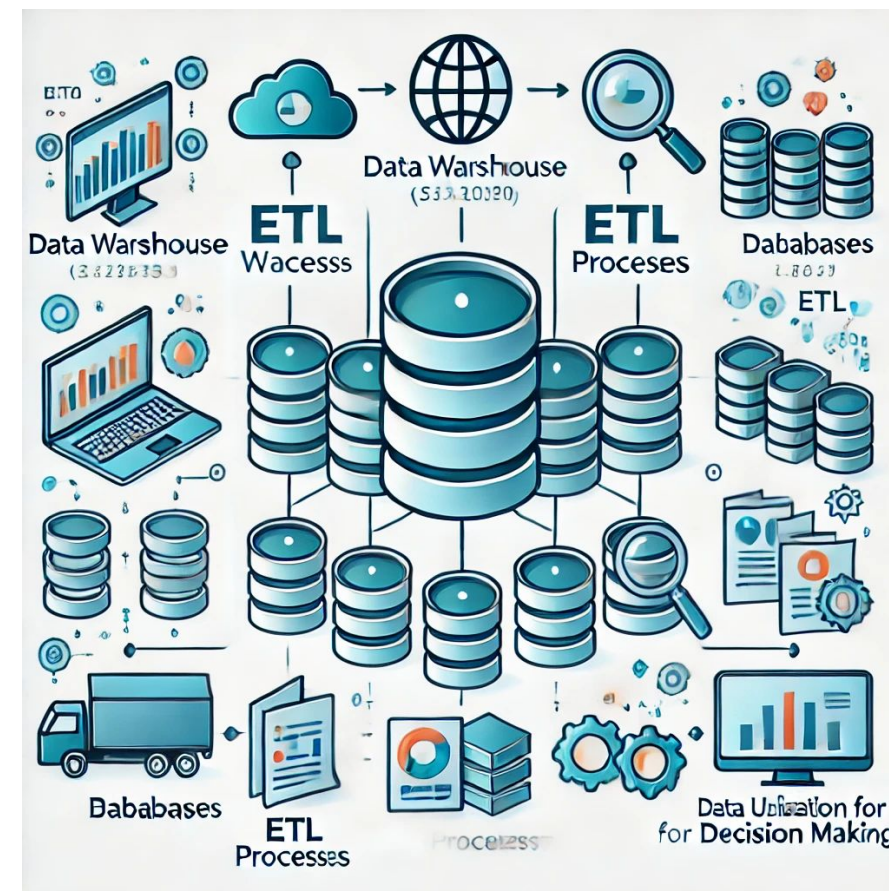
Bill H. Inmon

- “Una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis” o “la unión de todos los Data marts de una entidad”

Ralph Kimball

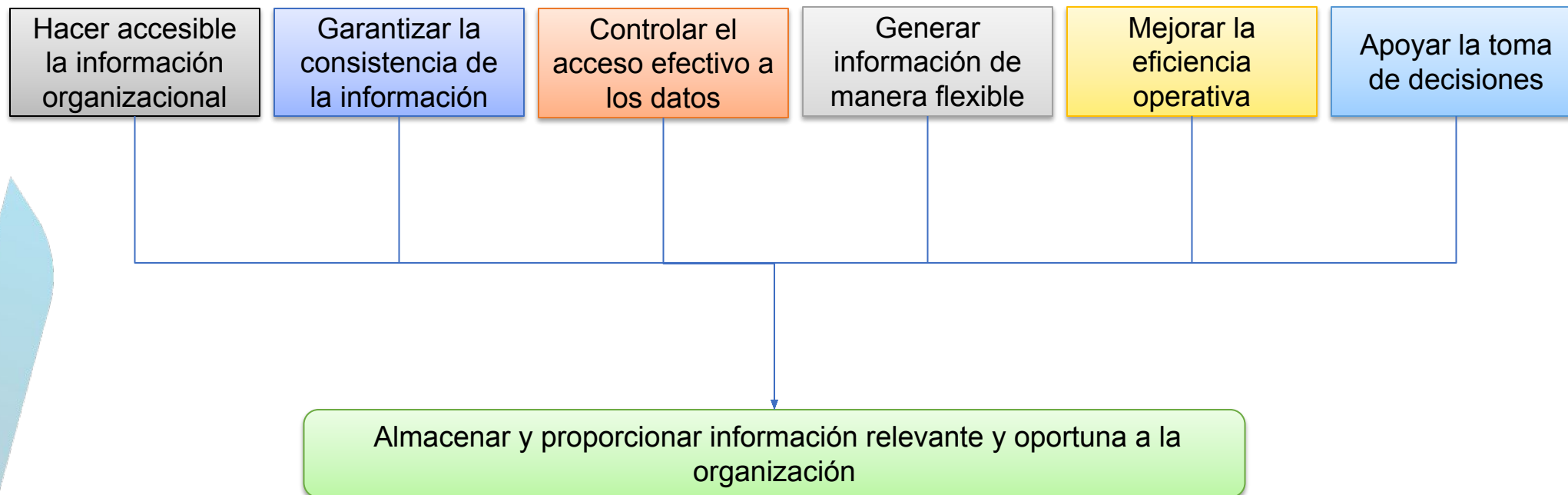
- “Yo considero al DW como algo que provee dos beneficios empresariales reales: Integración y Acceso de datos. DW elimina una gran cantidad de datos inútiles y no deseados, como también el procesamiento desde el ambiente operacional clásico”

Susan Osterfeldt





Objetivos de Data Warehouse



¿Cómo es un Data Warehouse?



Características

- Tomando como referencia la definición de Inmon





Características: Orientada al Negocio

- La información se clasifica según los aspectos que son de interés para la organización
- El diseño y la implementación del DW difieren significativamente de los procesos operacionales.
- Se excluye información que no se utilizará en el proceso de toma de decisiones.
- Los datos operacionales mantienen relaciones continuas entre tablas, mientras que en el DW las relaciones son múltiples e integradas.
- El DW se organiza alrededor de entidades de alto nivel, diseñadas para consultas e investigaciones.

Características: Integrada

- Datos de diversas fuentes internas y externas se consolidan antes de ser agregados al DW
- Incluye técnicas como ETL (Extracción, Transformación y Carga).
- Aborda problemas relacionados con nombres, unidades de medida, y codificaciones.
- Los datos son analizados y limpiados para asegurar su calidad antes de la carga.
- Los datos se almacenan en un modelo único y consistente, facilitando su uso por los usuarios finales.



Características: Variante en el tiempo

- Todos los datos en el DW tienen un sello de tiempo
- Permite el desarrollo de análisis de tendencias y patrones
- Los datos se procesan como series de instantáneas, representando diferentes periodos de tiempo.
- Es crucial definir la granularidad y periodicidad de los datos para evitar desbordamientos.
- Facilita el acceso a diferentes versiones de la misma información para análisis comparativos.

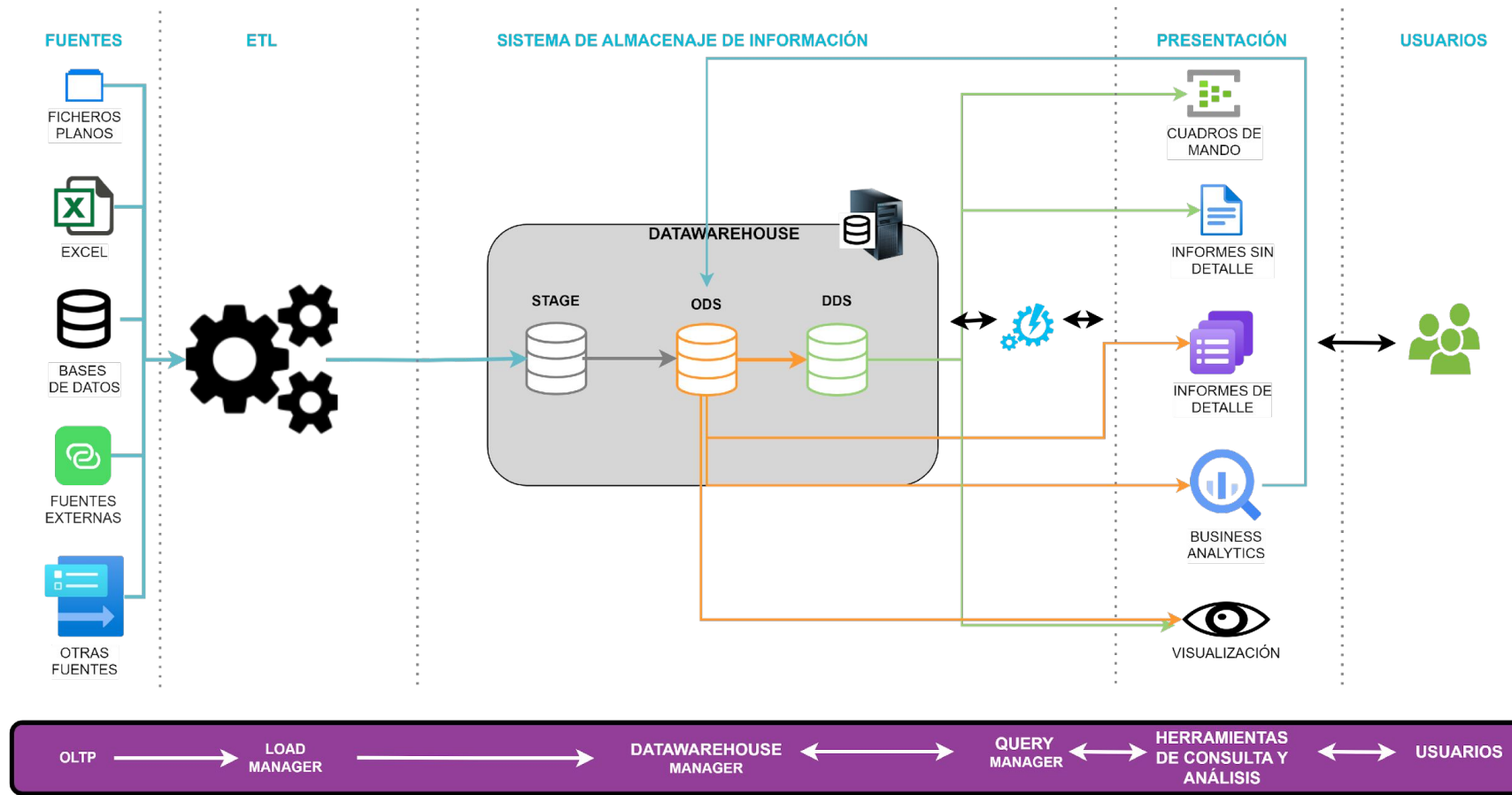


Características: No volátil

- Los datos no cambian una vez ingresados en el DW.
- Las operaciones básicas son la carga y el acceso a datos.
- No se necesitan mecanismos de control de concurrencia y recuperación.
- En el ambiente operacional, los datos varían constantemente, mientras que en el DW permanecen estables.
- La estabilidad de los datos hace que sean útiles para el análisis y la toma de decisiones.

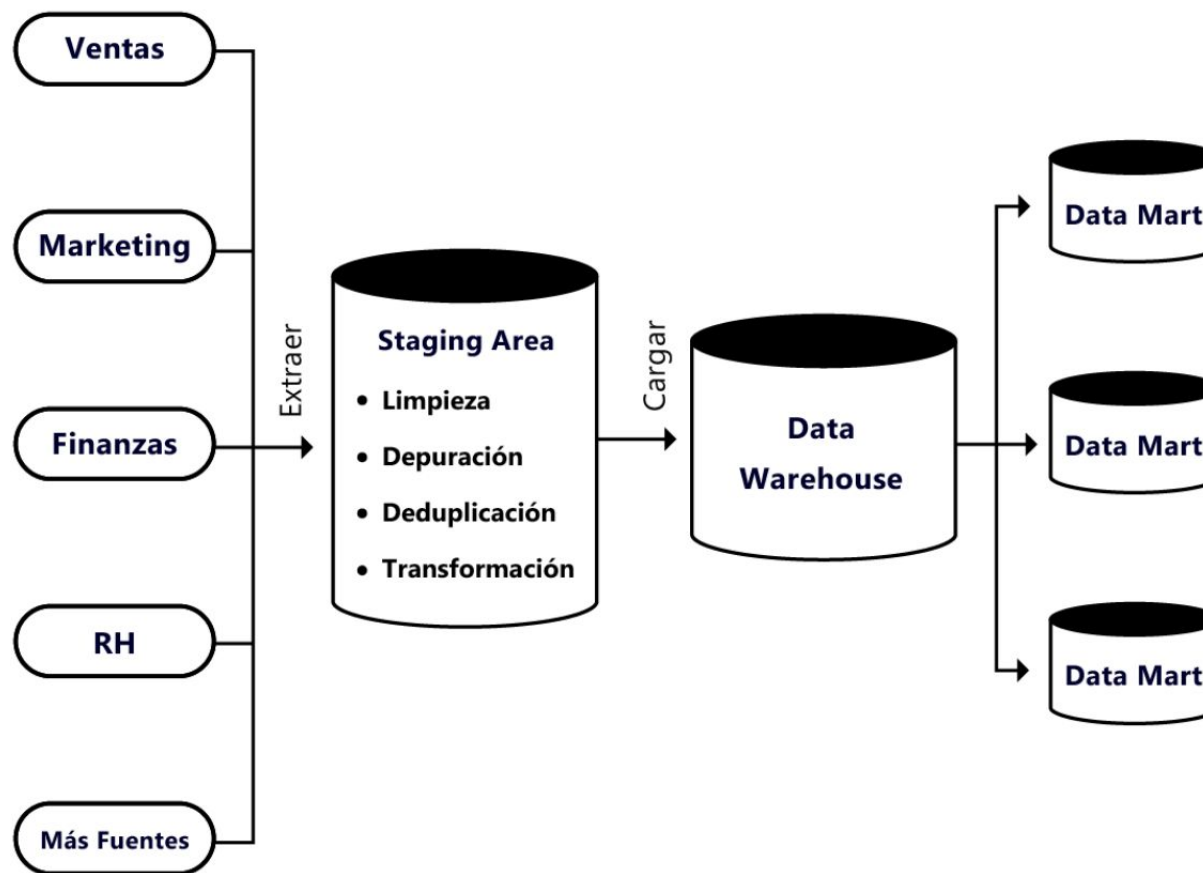


Arquitectura de un Data Warehouse



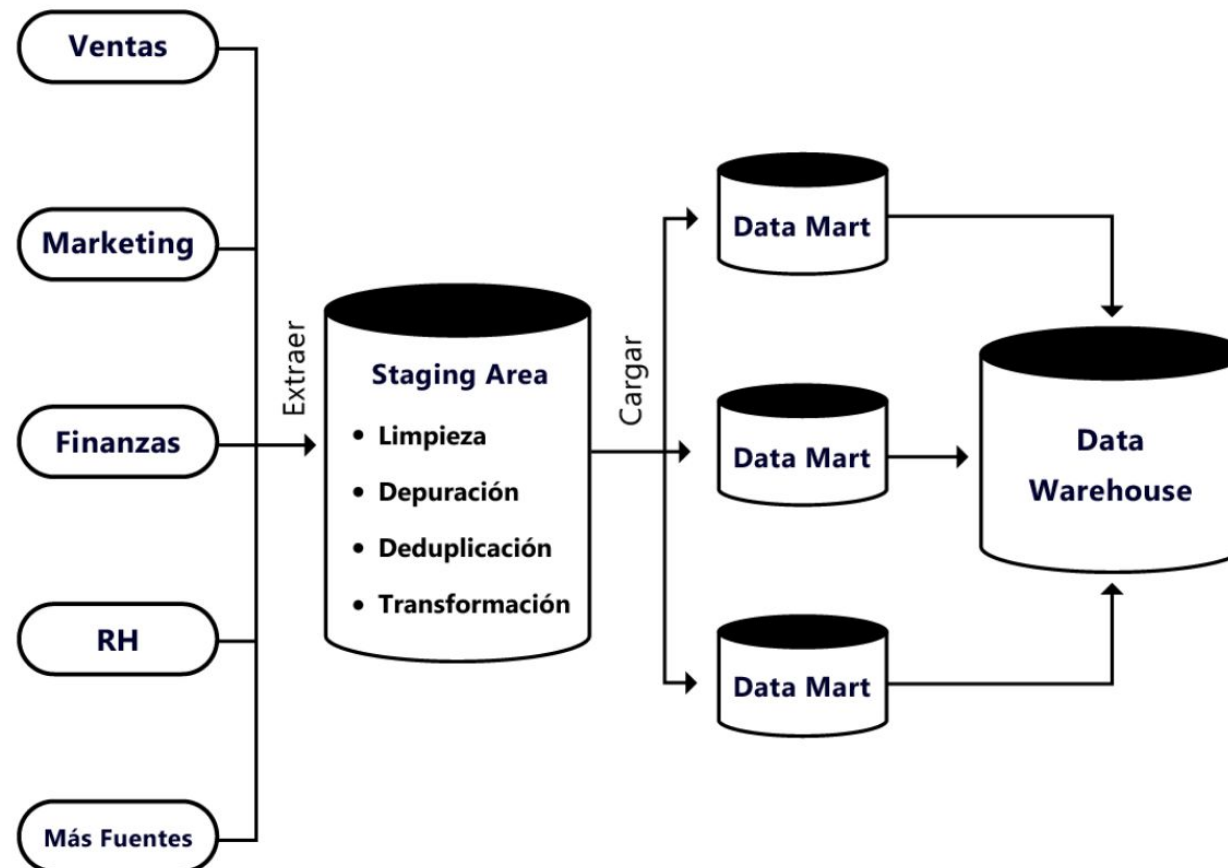
¿Cómo se construye un Data Warehouse?

- Enfoque de Data Warehouse Corporativo
 - Top-down
 - Bill Inmon



¿Cómo se construye un Data Warehouse?

- Enfoque de Data Mart
 - Bottom-up
 - Ralph Kimball



Fuentes de Datos

- **Fuentes de datos dispares (heterogéneos):** Debido al formato inconsistente de los datos, es común encontrar múltiples versiones de la misma información, lo que genera redundancia. El proceso ETL (Extracción, Transformación y Carga) ayuda a solucionar este problema al integrar y refinar estos datos heterogéneos en un entorno de trabajo antes de cargarlos en el Data Warehouse.
- **Fuente de datos interna:** Consiste en la recopilación de datos provenientes de usuarios individuales dentro de la organización. Estos datos internos aumentan la complejidad de la transformación e integración debido a la diversidad de fuentes, como hojas de cálculo y documentos de texto.

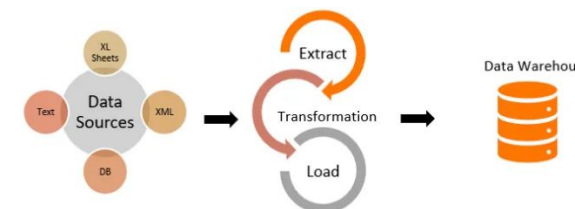
Fuentes de Datos

- **Fuente de Datos Externa:** Esta fuente utiliza datos recolectados de fuentes externas, como estadísticas de la industria y cuotas de mercado de competidores. Estos datos permiten a la organización detectar tendencias y comparar su rendimiento con el de sus competidores. Los datos externos deben ser convertidos a formatos apropiados antes de ser utilizados en la organización.
- **Metadatos:** Los metadatos definen el contenido del Data Warehouse e incluyen información sobre los procesos, fuentes de datos, almacenamiento, seguridad y autenticación del sistema.



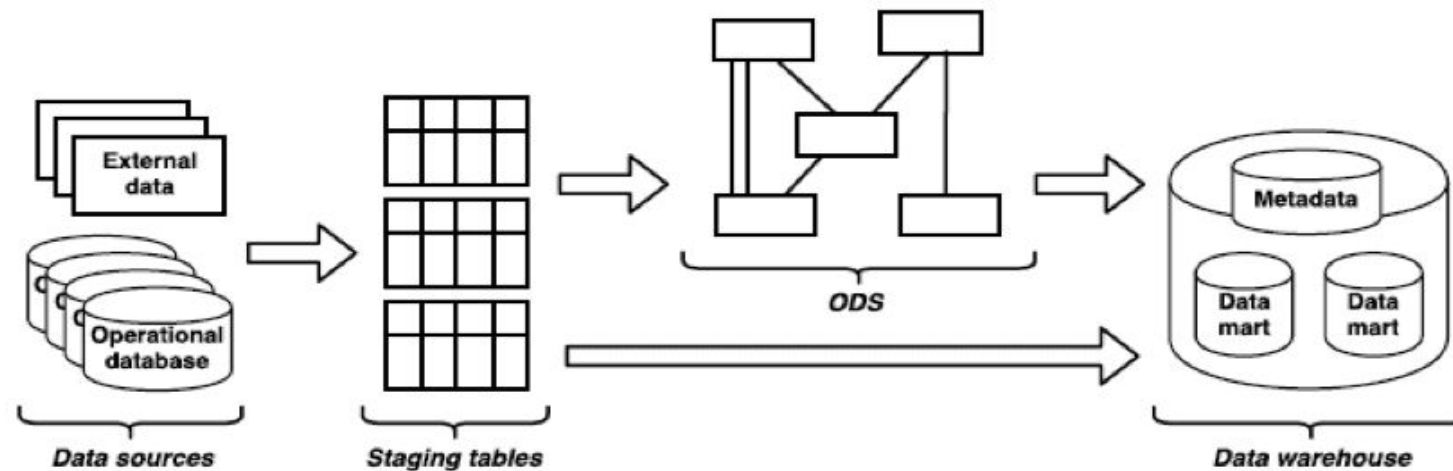
ETL

- **Extracción de datos:** Es el primer paso en el proceso de ETL. Este paso consiste en obtener información de diversas fuentes de origen, tanto internas como externas. Durante la extracción, se identifican los datos deseados y se extraen de múltiples fuentes diferentes, incluidos sistemas de bases de datos y aplicaciones. Una vez que los datos se han extraído, deben ser transportados físicamente al sistema de destino o a un sistema intermedio para su posterior procesamiento y/o transformación.
- **Transformación:** Es el siguiente paso, donde los datos se filtran, limpian, depuran, homogeneizan y agrupan. Este paso incluye la agregación de datos de diferentes fuentes. La transformación se lleva a cabo mediante el uso de reglas o tablas de consulta o mediante la combinación de datos con otros datos.
- **Carga:** Es el proceso de escribir los datos en el Data Warehouse. La fase de carga implica el momento en que los datos transformados se cargan en el sistema de destino.



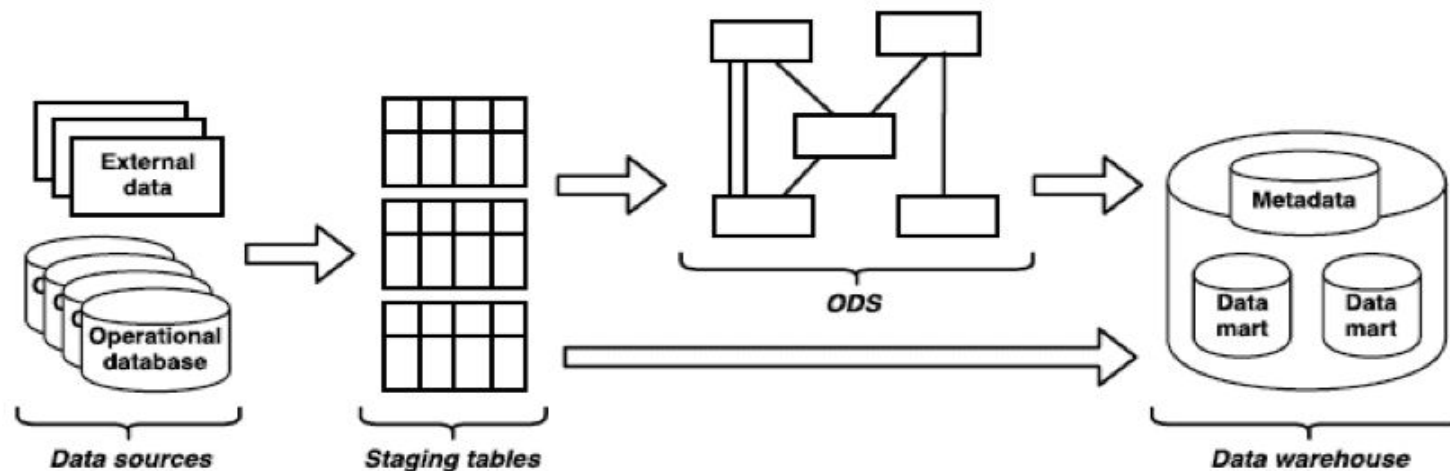
Stage (Área de preparación)

Es el sistema que se encuentra entre las fuentes de datos y el Data Warehouse. Es una parte esencial de la arquitectura de un Data Warehouse y su propósito es mejorar la calidad de los datos que ingresan al DWH. Esto se logra a través de un procesamiento previo que los datos reciben en esta etapa.



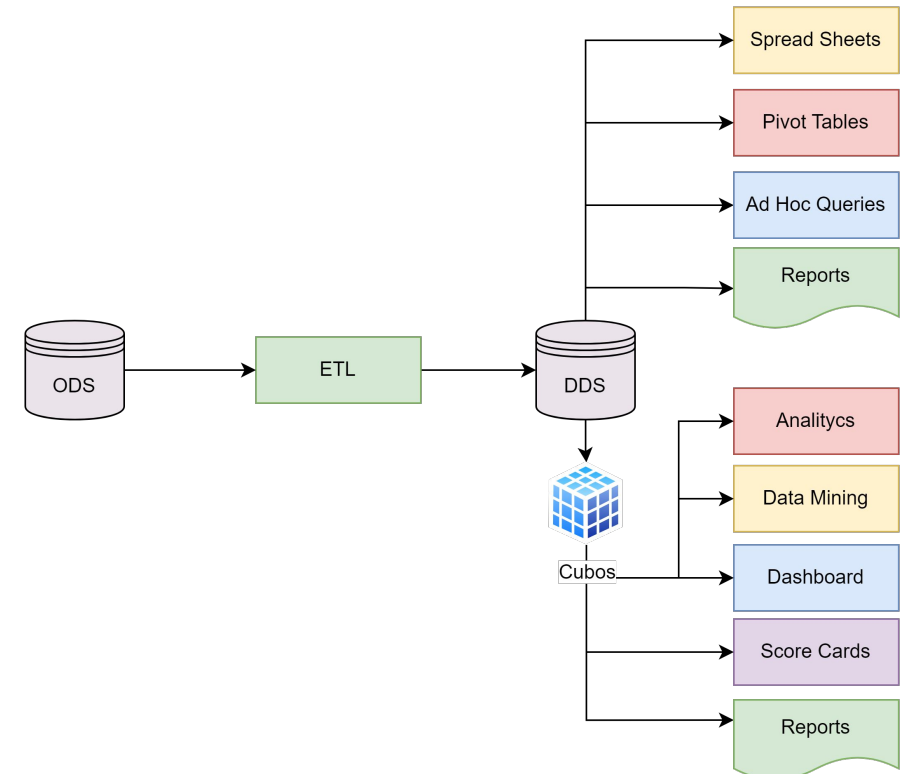
ODS (Almacén de datos operativos)

Es un área que da soporte a los sistemas operacionales mediante una estructura relacional y normalizada, permitiendo consultas y reportes. Aunque no almacena datos históricos, muestra una instantánea actualizada periódicamente desde la staging area, donde se limpian y transforman los datos. El ODS asegura la calidad e integridad de los datos, adaptándolos a las necesidades del negocio para su posterior análisis en el Data Warehouse.



DDS (Almacén de datos dimensionales)

Un DDS (Dimensional Data Store) es un almacén de datos orientado al usuario, estructurado en una o más bases de datos relacionales donde los datos están organizados en un formato dimensional. Esto facilita la realización de consultas analíticas. Los DDS a menudo utilizan cubos OLAP (Online Analytical Processing) para permitir un análisis multidimensional eficiente de grandes volúmenes de datos.



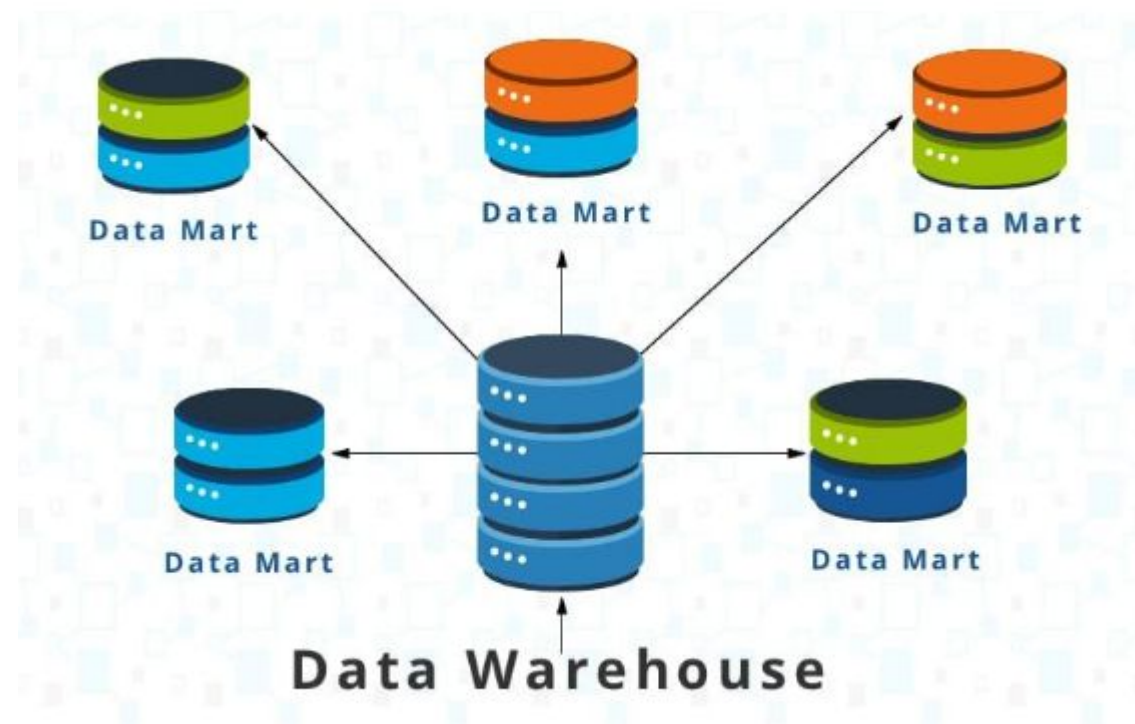
Datamart

Un data mart es una base de datos enfocada en una unidad de negocio específica dentro de una organización.

Se especializa en almacenar información relevante para un área particular de la empresa. Este puede obtener sus datos directamente de un data warehouse o compilar información de diversas fuentes externas e internas por sí mismo.



Universidad
Nacional de
Cajamarca
"Norte de la Universidad Peruana"





Datamart vs Data Warehouse

Data Warehouse

Un data warehouse se construye para funcionar como el repositorio central de datos de toda la empresa.

Datamart

Un data mart se diseña para satisfacer las necesidades específicas de una división o departamento particular de la empresa.



Tipos de Datamart



Data Mart dependiente: Se construyen utilizando un almacén de datos corporativo ya existente.



Data Mart independiente: Se crean sin la necesidad de un almacén de datos preexistente. Se crea a partir de fuentes externas



Data Mart Híbridos: Combina datos de un almacén de datos existente con datos provenientes de otros sistemas operativos

Ventajas de Datamart



**Universidad
Nacional de
Cajamarca**
"Norte de la Universidad Peruana"

- Agiliza las consultas disminuyendo la cantidad de datos que se deben revisar.
- Organiza los datos para facilitar su acceso a través de herramientas específicas.
- Distribuye los datos en diferentes plataformas de hardware.
- Reducción de costos ya que construir un Data Mart es mucho más económico que desarrollar un Data Warehouse.



Dashboard

Son herramientas visuales que permiten a los usuarios interactuar con los datos de manera intuitiva y comprensible. Estos dashboards se utilizan para presentar la información procesada y almacenada en el Data Warehouse de una manera gráfica y dinámica, facilitando la toma de decisiones basadas en datos.





Balance Scorecard

Es un modelo de gestión estratégico - Operacional que permite a las organizaciones traducir su visión y estrategia en un conjunto coherente de indicadores de desempeño.

Proporcionan una visión balanceada del rendimiento organizacional, ayudando a monitorear y ejecutar la estrategia empresarial.

Perspectives

Components

	Objectives	Measures	Targets	Initiatives
Financial	Increase Revenue	Current Budget	Increase by \$50,000	<ul style="list-style-type: none">Improve servicesBe more proactive with grant seeking
Customer	Offer More Analysis Services	Number of Analyses Offered	Increase by 10%	<ul style="list-style-type: none">Perform review of staff skill set
Internal Business Processes	Increase Applications for Grants	Number of Grants Applied for	10 per year	<ul style="list-style-type: none">Develop Proposal process
Learning & Growth	Monitor for New Grants	Number of Websites Searched	12 per week	<ul style="list-style-type: none">Teach Junior Staff to SearchDevelop resource sheet



Balance Scorecard

Se compone de cuatro perspectivas clave que permiten a las organizaciones medir y gestionar su desempeño integralmente:

- 1.Perspectiva Financiera:** Se centra en los objetivos financieros y las medidas de rendimiento económico que indican si la estrategia está contribuyendo a la mejora de los resultados financieros.
- 2.Perspectiva del Cliente:** Se enfoca en los objetivos y métricas relacionadas con la satisfacción, retención y adquisición de clientes, así como la cuota de mercado y la imagen de la empresa.
- 3.Perspectiva de los Procesos Internos:** Evalúa la eficiencia y eficacia de los procesos internos críticos para el éxito de la estrategia. Incluye medidas de productividad, calidad y ciclo de vida de los productos.
- 4.Perspectiva de Aprendizaje y Crecimiento:** Mide la capacidad de la organización para innovar, mejorar y aprender. Incluye indicadores relacionados con la formación y desarrollo de los empleados, la cultura organizacional y la infraestructura tecnológica.



OLTP (On Line Transaction Processing)

Es un sistema de gestión de bases de datos que se utiliza para sistemas transaccionales y están optimizada para registrar y gestionar transacciones diarias en tiempo real.

Estas bases de datos están diseñadas para el acceso rápido y la recuperación de datos, lo que las hace ideales para procesar transacciones a alta velocidad. Los sistemas OLTP permiten a las organizaciones manejar eficientemente un gran volumen de transacciones, asegurando rapidez y eficacia en su procesamiento.



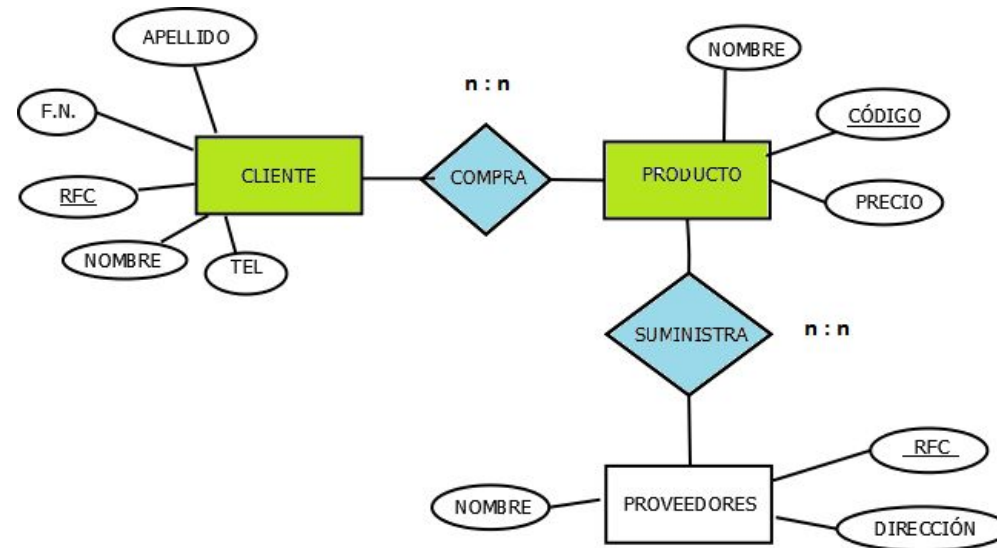
Características de OLTP

- **Procesamiento rápido de transacciones:** Optimizado para manejar grandes cantidades de transacciones de manera eficiente y rápida.
- **Uso en sistemas transaccionales:** Ideal para aplicaciones que requieren acceso y recuperación de datos rápidos para procesar volúmenes altos de pequeñas transacciones.
- **Alto volumen de pequeñas transacciones:** Diseñado para manejar inserciones, actualizaciones y eliminaciones de pequeñas cantidades de datos.
- **Estructura de datos normalizada:** Minimiza la redundancia y asegura la coherencia de los datos, permitiendo un almacenamiento y recuperación eficiente.
- **Optimizado para operaciones de escritura:** Eficiente en inserciones, actualizaciones y eliminaciones en tiempo real.
- **Baja latencia de datos:** Proporciona actualizaciones en tiempo real, asegurando la disponibilidad inmediata de datos actualizados.

Diseño de un sistema OLTP

- Para diseñar un sistema OLTP, se utiliza un Diagrama Entidad-Relación (DER). Este diagrama es una representación gráfica que captura la realidad mediante:

- Entidades
- Atributos
- Relaciones



En el diagrama, cada entidad y su relación se ilustran para facilitar la comprensión del sistema y su estructura de datos.

Ejemplos de OLTP

- Sistemas de comercio electrónico
- Sistemas de punto de venta
- Sistemas de reserva de vuelos
- Sistemas de atención médica
- Sistemas de reserva de hoteles
- Sistemas bancarios





OLAP (On Line Analytical Processing)

Es un sistema de gestión de bases de datos utilizado para sistemas analíticos. Estas bases de datos están optimizadas para realizar consultas complejas, adaptándose a sistemas que requieren procesar grandes volúmenes de datos para análisis y generación de informes.

Forma parte de lo que se conoce como sistemas analíticos, esta formada por un motor y un visor

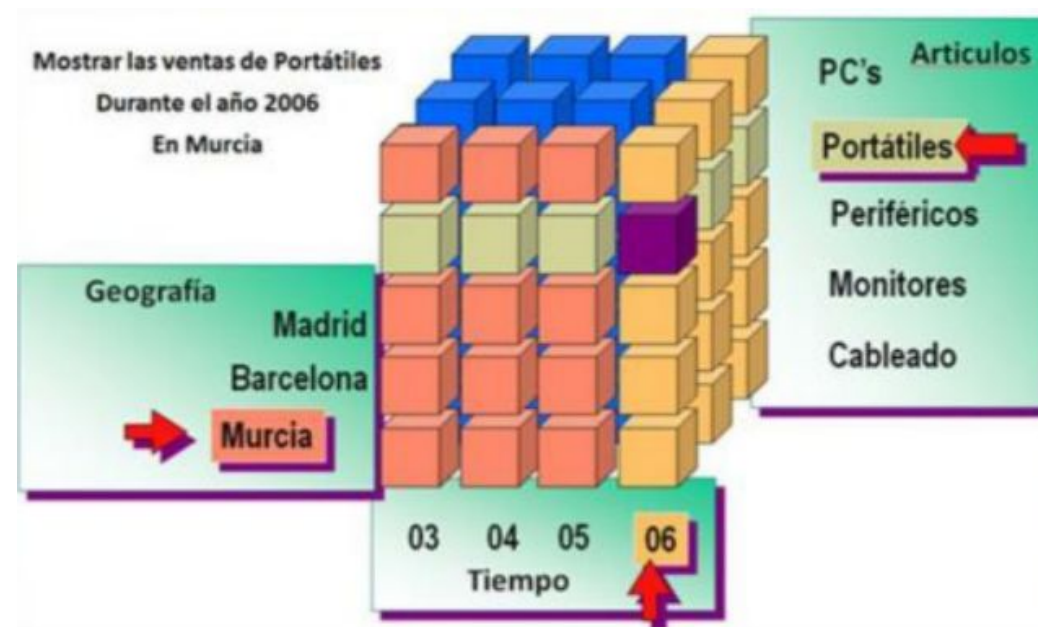


Características de OLAP

- **Enfoque en consultas complejas:** Las bases de datos OLAP están diseñadas para manejar consultas que implican múltiples dimensiones y jerarquías, facilitando el análisis avanzado de datos.
- **Análisis multidimensional:** Optimizado para analizar datos a lo largo de varios ejes, permitiendo explorar relaciones y correlaciones.
- **Uso en sistemas analíticos:** Comúnmente utilizado en herramientas de BI, almacenamiento de datos y sistemas de soporte a decisiones.
- **Bajo volumen de grandes transacciones:** Manejan transacciones grandes pero en bajo volumen, enfocándose en el análisis de datos.
- **Estructura de datos desnormalizada:** Almacenan datos de forma que se minimizan las uniones complejas, mejorando el rendimiento.
- **Optimizado para operaciones de lectura:** Diseñado para manejar grandes cantidades de consultas de recuperación de datos.
- **Alta latencia de datos:** Hay un retraso en la disponibilidad de los datos debido al procesamiento y agregación necesarios antes del análisis.

Configuración de OLAP

- Un cubo OLAP es un objeto que contiene medidas y dimensiones relacionadas, configurado en una base de datos de Analysis Services. Se pueden definir múltiples cubos dentro de una única base de datos, usando las mismas dimensiones o diferentes. También es posible tener un único cubo con varios grupos de medida en la misma base de datos, en lugar de definir cubos independientes.



Ejemplos de OLAP

- **Sistemas de Business Intelligence (BI):** estos sistemas permiten a las organizaciones analizar y visualizar datos de diversas fuentes para obtener información sobre el rendimiento empresarial, identificar tendencias y tomar decisiones basadas en datos.
- **Sistemas de almacenamiento de datos:** estos sistemas almacenan grandes volúmenes de datos de diversas fuentes y ofrecen una vista unificada de los datos para fines de análisis. Sirven como un repositorio centralizado, lo que permite a las organizaciones acceder y analizar datos de múltiples fuentes de manera optimizada.
- **Sistemas de análisis financiero:** estos sistemas permiten a los analistas financieros realizar análisis financieros complejos, como pronósticos, presupuestos y análisis de variaciones.
- **Sistemas de análisis de ventas:** estos sistemas permiten a los equipos de ventas analizar los datos de ventas por cliente, producto, región y otros parámetros para identificar tendencias y oportunidades de ventas.
- **Sistemas de análisis de marketing:** estos sistemas permiten a los equipos de marketing analizar el comportamiento del cliente, el rendimiento de la campaña y otras métricas de marketing para optimizar las estrategias de marketing.



Diferencias entre OLTP y OLAP

Sistema Operacional (OLTP)	Data Warehouse (Basado en Modelos Dimensionales: OLAP)
Almacena datos actuales.	Almacena datos históricos.
Almacena datos de detalle.	Almacena datos de detalle y datos agregados a distintos niveles.
Los datos son diámicos (actualizables).	Los datos son estáticos.
Las transacciones son repetitivas.	Los procesos no son previsibles.
El número de transacciones es elevado.	El número de transacciones es bajo o medio.
Dedicado al procesamiento de transacciones.	Dedicado al análisis de datos.
Orientado a los procesos de la organización.	Orientado a la información relevante.
Soporta decisiones diarias.	Soporta decisiones estratégicas.
Sirve a muchos usuarios administrativos.	Sirve a técnicos de dirección.



Universidad
Nacional de
Cajamarca
"Norte de la Universidad Peruana"

EVALUACIÓN DEL TEMA



Universidad Nacional de Cajamarca



www.unc.edu.pe/



Universidad Nacional de Cajamarca



REFLEXIONEMOS

- ¿Qué es un Data Warehouse y cuál es su propósito principal?
- ¿Cuáles son las características fundamentales de un Data Warehouse?
- ¿Cuál es la diferencia entre OLAP y OLTP?
- ¿Qué es un Datamart y cómo se diferencia de un Data Warehouse?
- ¿Qué son los cubos OLAP?
- ¿Cuáles son los procesos ETL y cuál es su importancia en la gestión de datos?
- ¿Qué papel juegan los Dashboards y los Balanced Scorecards en la arquitectura de un Data Warehouse?



Universidad
Nacional de
Cajamarca
"Norte de la Universidad Peruana"

ACTIVIDAD



Universidad Nacional de Cajamarca



www.unc.edu.pe/



Universidad Nacional de Cajamarca