



Programa PUCP DATA DRIVEN

Formación tecnológica

Dirección Académica de Planeamiento y Evaluación
Dirección de Gestión del Talento Humano





Profesor



Ing. Oscar Martinez M.Sc.

- Profesor asociado a Danalytics.
- Ingeniero Electrónico.
- Magister en Ingeniería Electrónica.
- Magister en Bioingeniería.

En la industria

- Ingeniero de Desarrollo de Software.
- Ingeniero en Modelos ML, Científico de Datos.

En la academia

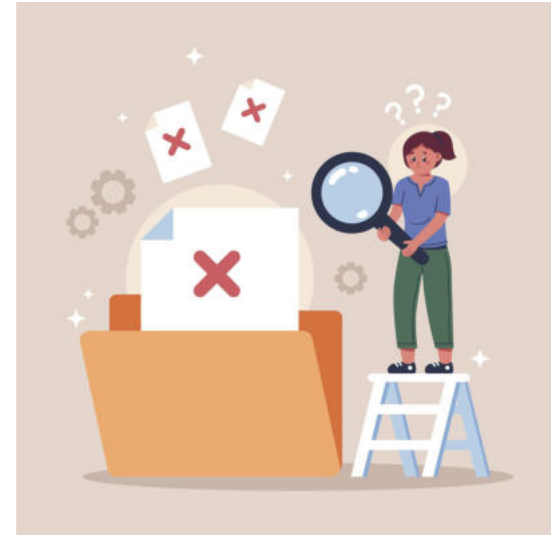
- Profesor Cátedra de Ing Electrónica e Ing de Sistemas.
- Director de la Especialización en IA (Modalidad Virtual) de la PUJ Bogotá.



Los participantes



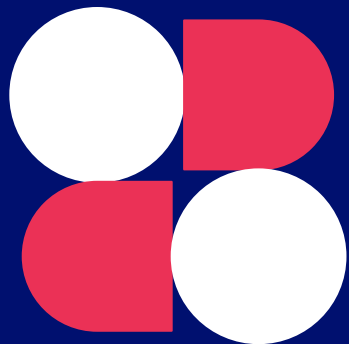
- ¿Alguna experiencia en programación? ¿SQL?
- ¿Algo puntual que espera y/o quiere aprender en este curso?



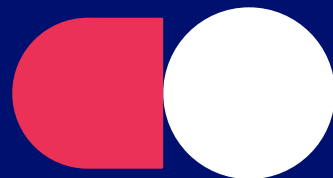


- **Sesiones Teórico prácticas 2 veces por semana**
- **Taller (ejercicios complementarios) 1 vez por semana**
 - Se comparten desde el miércoles y se desarrollan en la monitoria los viernes
 - Al finalizar la monitoria se comparten las respuestas a todos los participantes del curso
 - No obligatorio ni calificable pero altamente recomendable para afianzar conceptos



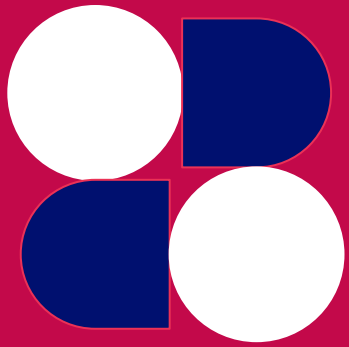


**1. Presentación del
profesor, interacción
con participantes y
reglas de juego**



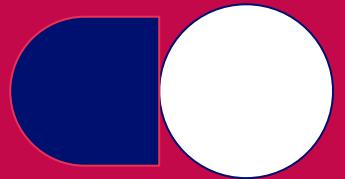


Tema	Contenido	Tiempo
Introducción a Bodega de Datos	• Conceptos básicos • Elementos de una bodega de datos • Diseño, implementación y buenas prácticas	4 horas
Arquitectura de ELT	• Conceptos básicos • Diferencias entre ELT y ETL • Buenas prácticas de ELT • Ejercicios prácticos de transformación de datos en Power Query de Power BI	6 horas
Introducción a SQL	• Fundamentos de SQL: Modelo Relacional, sintaxis básica, estructura de consultas. • Consultas simples: SELECT, FROM, WHERE, ORDER BY, Operaciones adicionales. • Operaciones de manipulación de datos: SUM, AVG, MIN, MAX, HAVING • Unión de datos: JOIN, Consultas anidadas • Ejercicios prácticos para aplicar SQL en la extracción y manipulación de datos.	8 horas



SESIÓN 1:

Introducción a Data Warehousing





Al finalizar la sesión, lograrás:

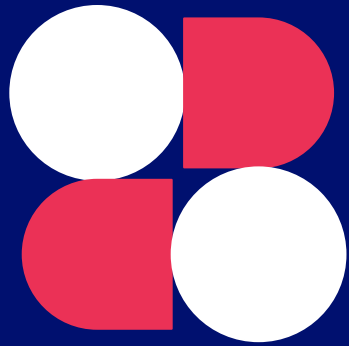
- Comprender los Conceptos Básicos de las Bodegas de Datos
- Identificar los elementos que componen una Bodega de Datos y los diferencian de bases de datos relacionales
- Comprender los servicios derivados de una Bodega de Datos



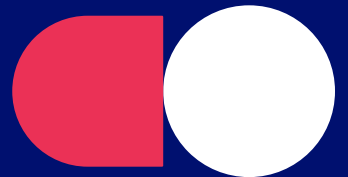
Contenido:



- 1 ¿Qué es una bodega de datos?
- 2 Conceptos fundamentales de una bodega de datos
- 3 Diferencias entre bodega de datos y base de datos relacional
- 4 Roles clave en gestión y aplicación
- 5 Servicios y Elementos Derivados de las Bodegas de Datos
- 6 Aplicación de las bodegas de datos para la PUCP



1. ¿Qué es una bodega de datos?



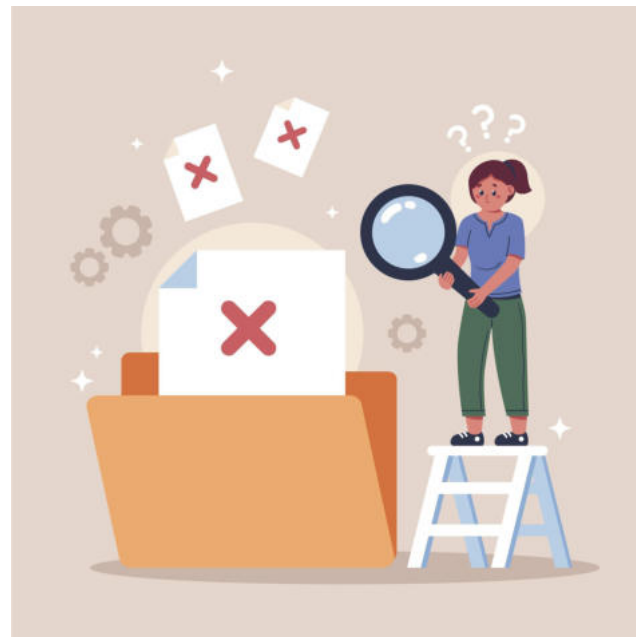


I. ¿Qué es una bodega de datos?



Situación actual: Desafíos en las universidades

- Producción y recopilación de datos a un ritmo sin precedentes
- Falta de análisis sobre los [datos](#)





I. ¿Qué es una bodega de datos?

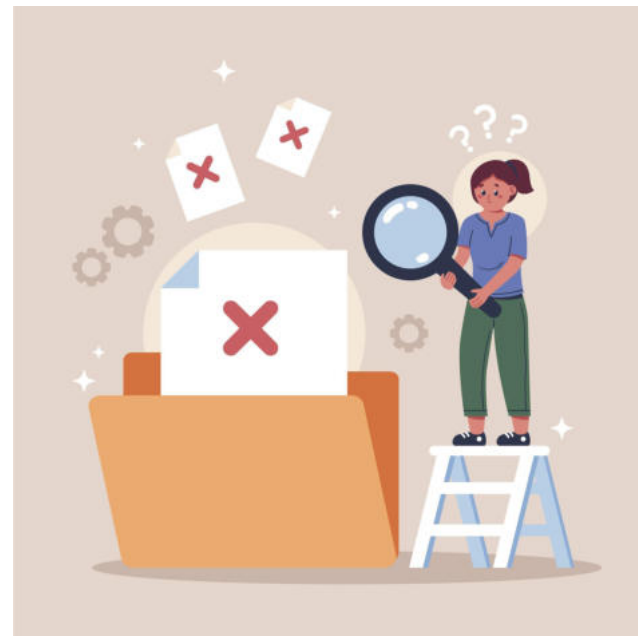


Situación actual: Desafíos en las universidades

- Producción y recopilación de datos a un ritmo sin precedentes
- Falta de análisis sobre los datos

Retos

- Dispersión de los datos
- Problemas en la calidad de datos





I. ¿Qué es una bodega de datos?



Sistema de almacenamiento de datos diseñado para:

- Recopilar
- Organizar
- Consolidar
- Optimizar

Grandes volúmenes de información provenientes de diversas fuentes dentro de una organización

Para:

- Centralizar
- Asegurar coherencia en los datos

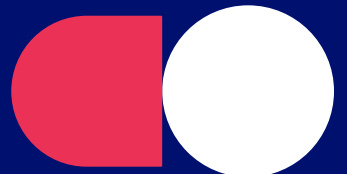
Con el objetivo final de:

- Analizar
- Apoyar la toma de decisiones estratégicas





2. Conceptos fundamentales de una bodega de datos





2. Conceptos fundamentales de una BdD



ELT - ETL (Extract, Transform, Load)

- Extracción - Carga - Transformación
- Extracción - Transformación - Carga
- Producción y recopilación de datos a un ritmo sin precedentes
- Falta de análisis sobre los datos

Modelos dimensionales (Ya lo conocemos!)

- Tablas de Dimensiones
- Tablas de Hechos
- Optimización de lectura

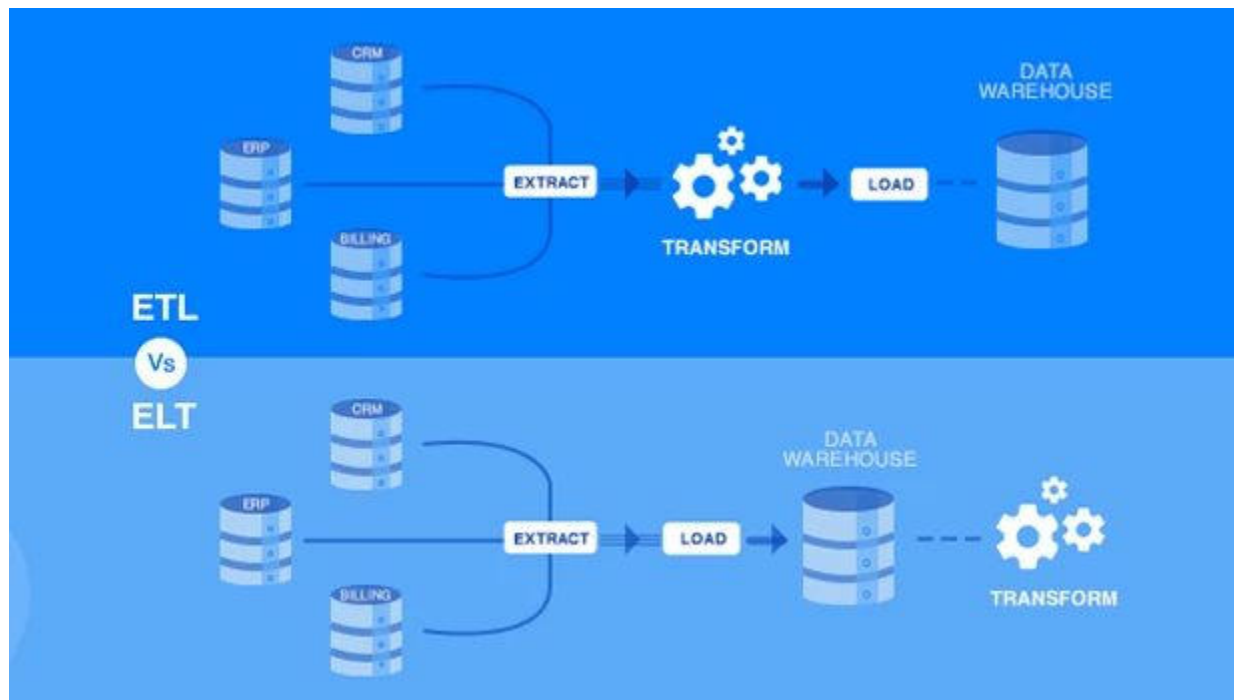
Consultas OLAP (Ya lo conocemos!)

- Dimensiones
- Hechos





2. ELT vs ETL (Sesión 4!)





shutterstock.com · 142837480

slideswith.com/CIREPASOS1



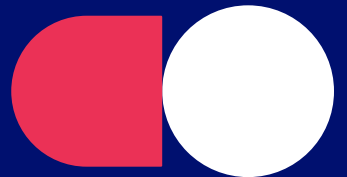
2. Conceptos fundamentales de una BdD



- I. El sistema de gestión de inscripciones de la universidad registra cada vez que un estudiante se inscribe en un curso. ¿Qué tipo de sistema es este?
R// B. OLTP
- I. El departamento de análisis académico de la universidad quiere analizar las calificaciones de los estudiantes de los últimos diez años para identificar tendencias y mejorar la calidad educativa. ¿Qué tipo de sistema es más adecuado para este análisis?
R// A. OLAP
- I. ¿Qué tipo de tablas contienen datos transaccionales como ventas, transacciones o registros de eventos. ?
R// C. Tablas de hechos
- I. ¿Qué tipo de tablas contienen datos descriptivos relacionados con las tablas de hechos, como productos, clientes o fechas?
R// D. Tablas de dimensiones
- I. ¿Dentro de la PUCP se está utilizando ETL o ELT?
R//A. ELT



3. Diferencias entre bodega de datos y base de datos relacional





3. Diferencias entre BdD y base de datos relacional



Aspecto	Base de datos	Bodega de Datos
Estructura de datos	Almacena datos estructurados	Almacena datos estructurados y semiestructurados.
Volumen de datos	Diseñado para volúmenes de datos más pequeños	Creado para grandes volúmenes de datos históricos
Transformación de datos	Transformación mínima de datos	A menudo incluye procesos ETL (Extracción, Transformación, Carga)
Complejidad de la consulta	Adecuado para consultas transaccionales simples	Diseñado para consultas analíticas complejas
Uso	Utilizado para las operaciones del día a día.	Empleado para inteligencia empresarial, análisis e informes.
Estado latente	Proporciona acceso de baja latencia a los datos.	Puede tener una latencia más alta debido a consultas complejas
Cambios de esquema	Los cambios de esquema complejos pueden requerir tiempo de inactividad	Más flexibilidad para adaptarse a los cambios de esquema.
Retención de datos	Almacena datos recientes y que cambian con frecuencia	Conserva datos históricos para el análisis de tendencias.

¿Qué crees que falta?

¿Los datos en ambos siempre están actualizados?

En las bodegas de datos no es tan crítico que los datos estén 100% actualizados, ¿por qué?



3. Diferencias entre BdD y base de datos relacional



Imaginémonos lo siguiente:

- I. Suponga que desde la rectoría nos solicitan un reporte complejo (en términos de cálculos, indicadores y/o gráficos) asociado a la información que manejamos durante el día a día, ¿por qué podría ser mejor generarlo a partir de una bodega de datos?

- R// La bodega de datos nos ofrece:
 - Acceso a datos centralizados y confiables
 - Histórico de nuestros datos organizados
 - Potencial de automatización
 - Tranquilidad en el caso de mayores volúmenes de datos

2. ¿Les ha ocurrido que algún sistema dentro de la universidad se “pone lento”? ¿una bodega de datos podría apoyar esto de alguna manera?

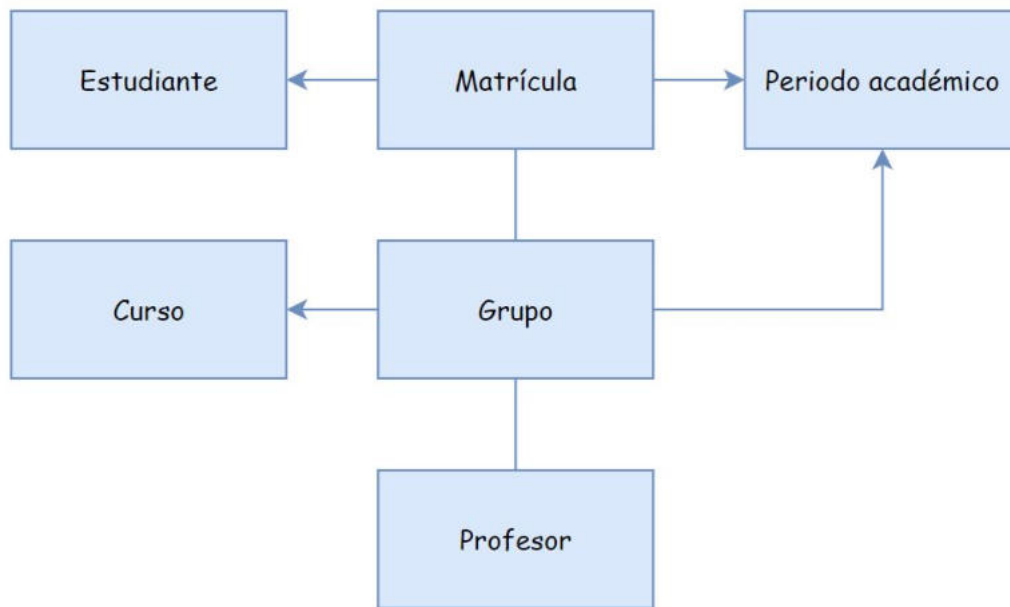
- R// Si claro, debido a que este tipo de sistemas (comúnmente) son tipo OLTP, motivo por el cual el uso de los estudiantes, profesores, administrativos y otros junto con consultas complejas pueden ocasionar sobrecarga, lentitud y mala experiencia de la comunidad universitaria



3. Diferencias entre BdD y base de datos relacional



Tiempos requeridos para la obtención de resultados



Número de registros por Tabla

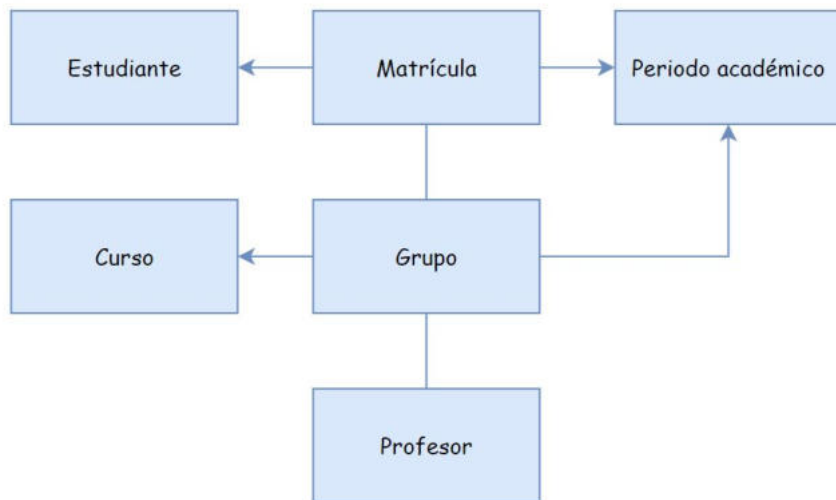
- Estudiante: 5,000 registros
- Periodo Académico: 10 registros
- Profesor: 200 registros
- Curso: 100 registros
- Grupo: 50 registros



3. Diferencias entre BdD y base de datos relacional



Tiempos requeridos para la obtención de resultados



Supongamos que queremos realizar una consulta para obtener la siguiente información:

El número de estudiantes matriculados en cada curso durante un período académico específico. Además, queremos obtener los nombres de los profesores que están dictando dichos cursos.

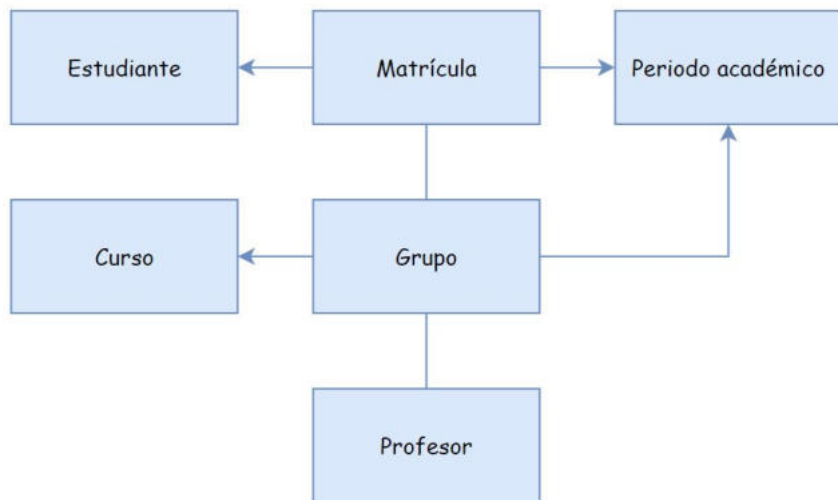
¿Cuántas operaciones de bases de datos (por ende tiempo) requieren hacerse para dar el resultado final?



3. Diferencias entre BdD y base de datos relacional



Tiempos requeridos para la obtención de resultados



Supongamos que queremos realizar una consulta para obtener la siguiente información:

El número de estudiantes matriculados en cada curso durante un período académico específico. Además, queremos obtener los nombres de los profesores que están dictando dichos cursos.

¿Cuántas operaciones de bases de datos (por ende tiempo) requieren hacerse para dar el resultado final?

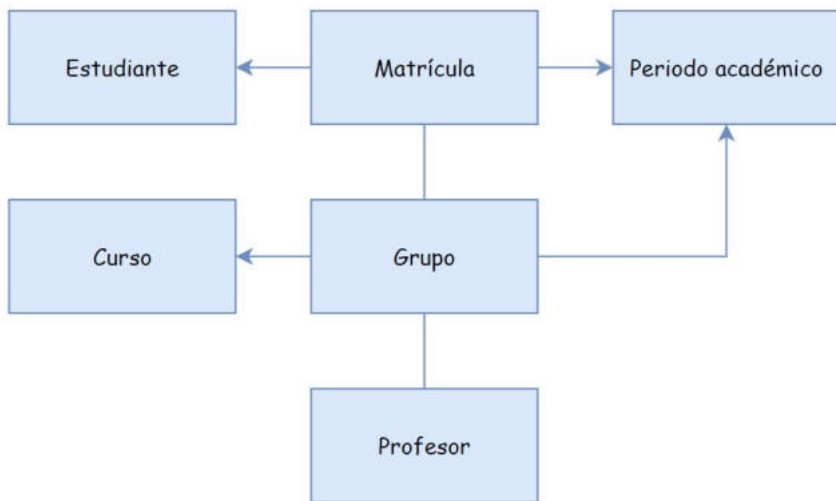
- Base de datos relacional: 15,000,000,000



3. Diferencias entre BdD y base de datos relacional



Tiempos requeridos para la obtención de resultados



Supongamos que queremos realizar una consulta para obtener la siguiente información:

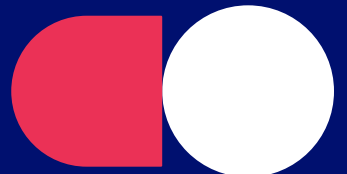
El número de estudiantes matriculados en cada curso durante un período académico específico. Además, queremos obtener los nombres de los profesores que están dictando dichos cursos.

¿Cuántas operaciones de bases de datos (por ende tiempo) requieren hacerse para dar el resultado final?

- Base de datos relacional: 15,000,000,000
- Bodega de datos: 758,000 (**99.49% menos**)



4. Roles clave en gestión y aplicación



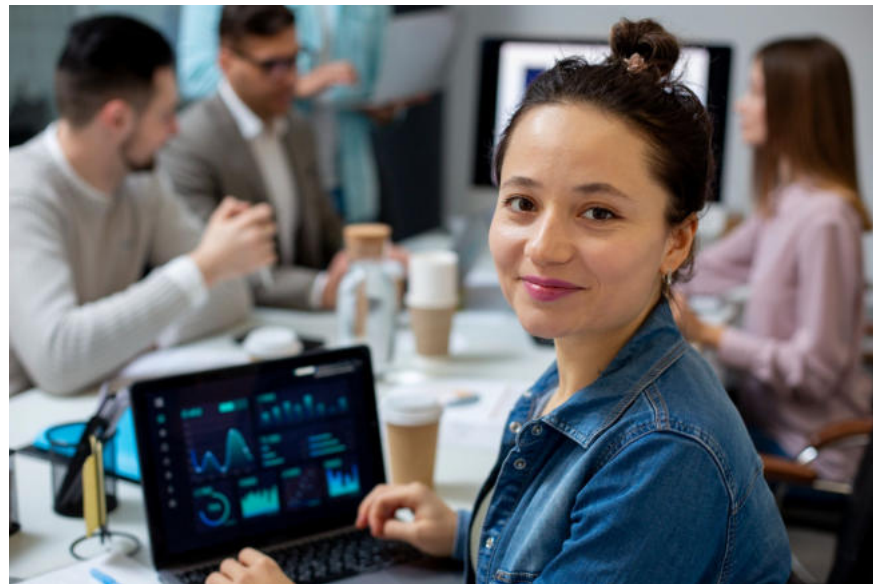


4. Roles clave en gestión y aplicación



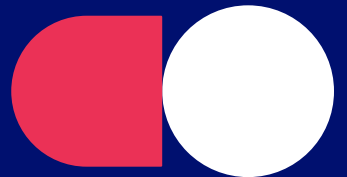
- **Administrador de la Bodega de Datos**
- **Analista de Datos**
- **Ingeniero de Datos**
- **Científico de Datos**

¿En el uso de la información qué rol tendríamos?





5. Servicios y Elementos Derivados de las Bodegas de Datos





5. Servicios y Elementos Derivados de las Bodegas de Datos



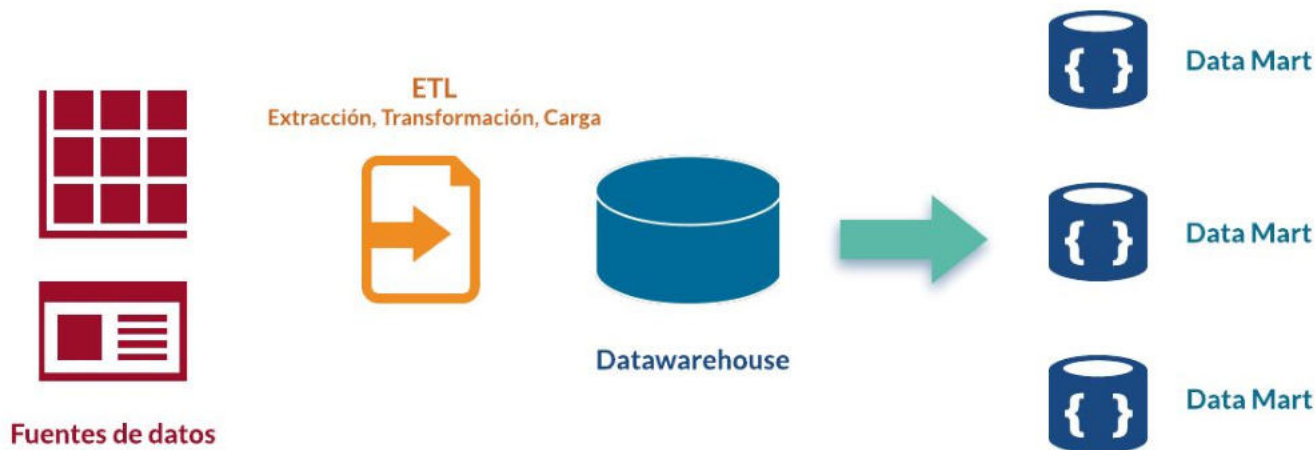
Preguntémonos:

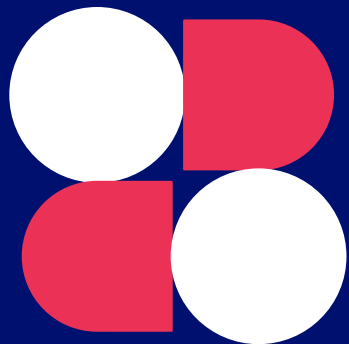
1. ¿Qué tanta información podría potencialmente existir en una bodega de datos?
2. ¿A todas las personas les interesa (o deberían) **TODA** la información disponible?
3. ¿Qué podríamos hacer?



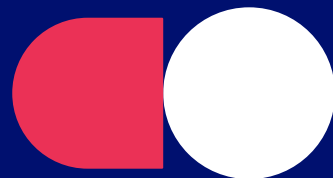


5. Servicios y Elementos Derivados de las Bodegas de Datos: Data Marts





6. Aplicación de las bodegas de datos para la PUCP





6. Ejercicio Práctico



- Grupos: 5 grupos
- Tiempo: 30-45 minutos

Descripción:

En grupos de 5 personas vamos a pensar (2) casos de aplicación de bodegas de datos:

1. ¿Qué tipo de aplicación de bodega de datos propondrías en tu área/dependencia? ¿por qué?
2. Recordemos que hay muchas áreas en nuestra universidad: ¿Qué tipo de aplicación de bodega de datos se podría originar que implique trabajar de la mano con otra área?

Y propongamos 2 preguntas que podrían ser de impacto para alguna dependencia (e.g. rectoría, vicerrectorías, facultades, extensión u otro) y explicar el porqué





Gracias

