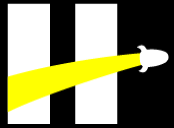


HENRY

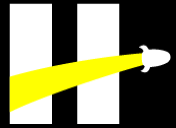
A bright yellow beam of light originates from the left edge of the frame and points towards the letter 'R' in the word 'HENRY'. The beam is wider on the left and tapers as it approaches the letter, which it appears to be illuminating or hitting.

Calidad del Dato



¿Para qué buscamos la calidad de los datos?

- La calidad de datos confiabiliza en análisis.
- Si utilizamos datos no confiables podemos llegar a conclusiones erróneas.
- Las fuentes son críticas en el proceso de selección de datos.
- Existen métodos de control, auditoría, corrección y confiabilización de datos.
- Recuperar información perdida o incompleta.
- Resolver conflictos en los datos.



Causas de la mala calidad de los datos

- Carga de datos en forma manual o Data Entry.
- Carga de datos externos sin los recaudos correctos para su adecuación.
- Problemas de carga originados en los sistemas transaccionales utilizados como fuente de datos.
- Implementación de nuevas aplicaciones en la organización, implica nuevos orígenes de datos, que necesitan ser congruentes con los datos ya existentes.
- Cambios en las aplicaciones existentes o migraciones de sus bases de datos.



Criterios

Actualización

- Los datos deben estar actualizados. Debe existir referencias de la fecha de confección o de la fecha de última actualización.
- Por ejemplo: Información de deuda sin una referencia en cuanto a la fecha de actualización.

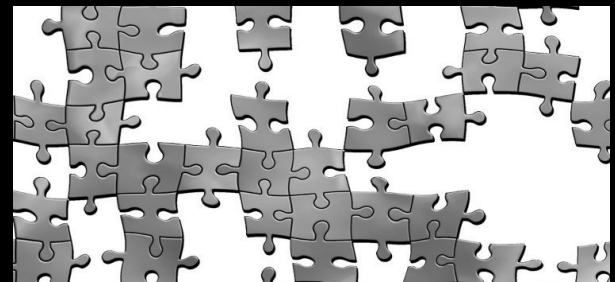




Criterios

Compleitud

- Los datos deben estar completos. Puede parecer obvio pero es una de las situaciones más habituales.
- Por Ejemplo: tablas con datos filiatorios y de contacto con campos vacíos aleatoriamente.



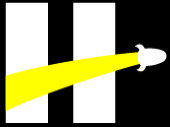


Criterios

Fiabilidad

- La procedencia y la trazabilidad del dato son características que hacen a la fiabilidad.
- Si trabajamos con una tabla de la que no conocemos su procedencia o bien que bajamos de una pagina poco confiable, los datos no serán fiables y nos pueden llevar a análisis erróneos.
- De igual manera, si no podemos reconstruir el camino completo del dato desde su captura hasta la actualidad, el conjunto de datos no tiene trazabilidad.



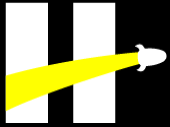


Criterios

Accesibilidad

- Los datos deben ser accesibles con bajo nivel de esfuerzo.
- Deben estar en lugares previsibles y ser fácilmente ubicables y elegibles.
- Ejemplo: Una tabla con nombres de campos numerados: Campo1, Campo2, etc...
- Ejemplo 2: Un reporte que se aloja en una ubicación poco habitual.



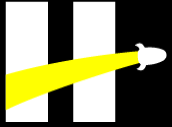


Criterios

Consistencia

- Interna: Calidad de caracteres y de lo que se guarda en los campos.
- Externa: Calidad de interdependencia y racionalidad de los campos.
- Ejemplo: Las claves primarias y las claves foráneas deben ser consistentes y permitir la relación entre tablas.





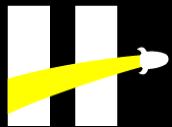
Preparación de los Datos

- Integración
- Limpieza del ruido
- Transformación
- Imputación de valores faltantes.

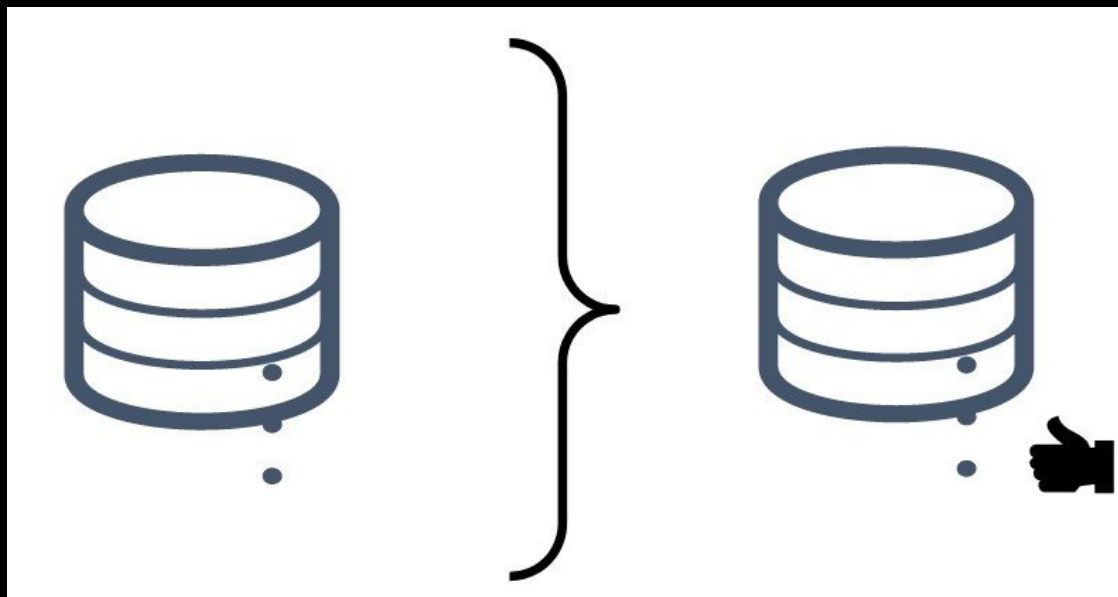


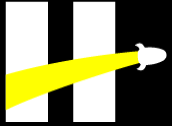
Integración



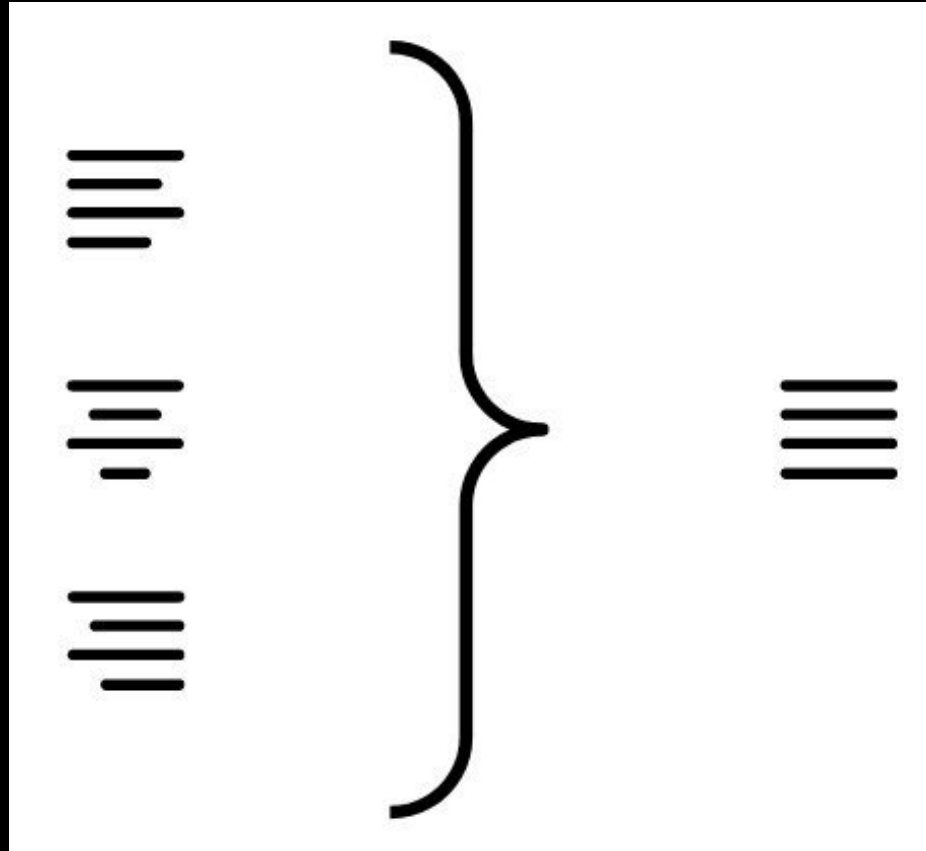


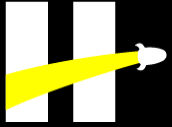
Limpieza



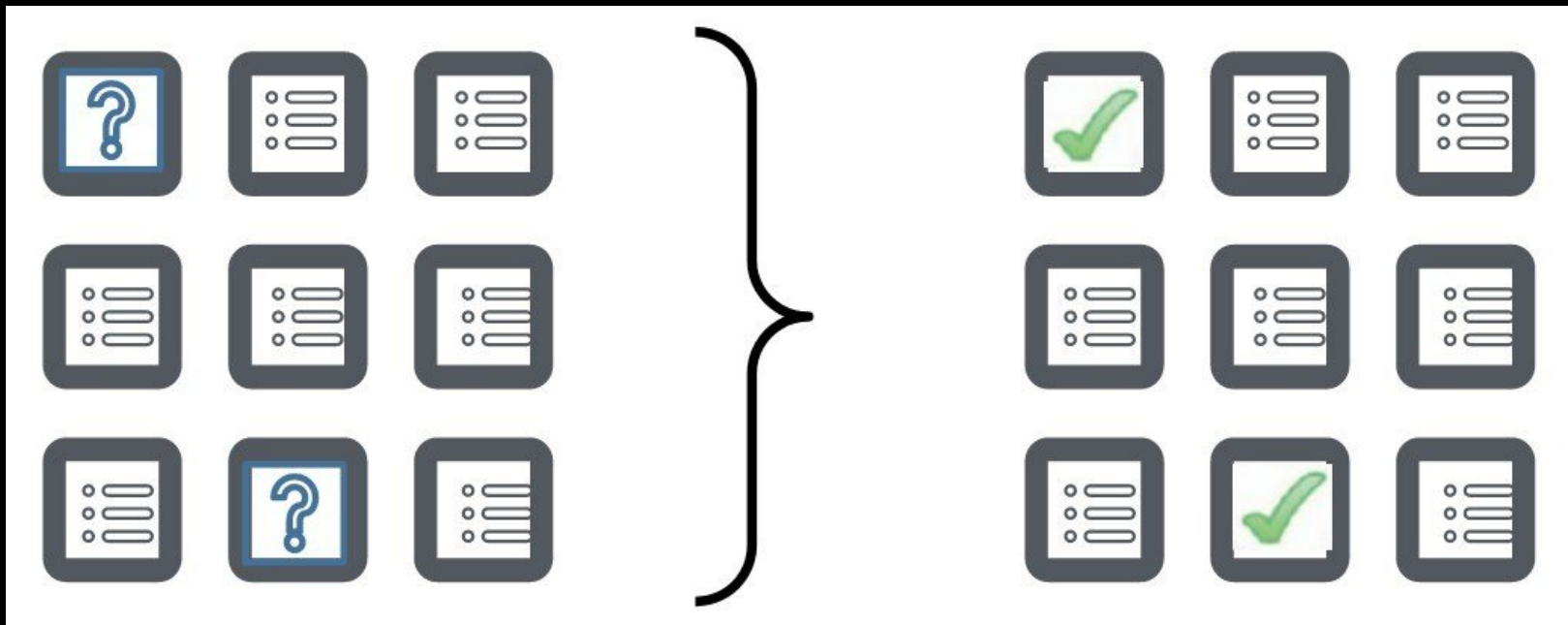


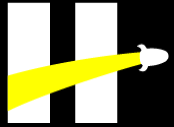
Normalización



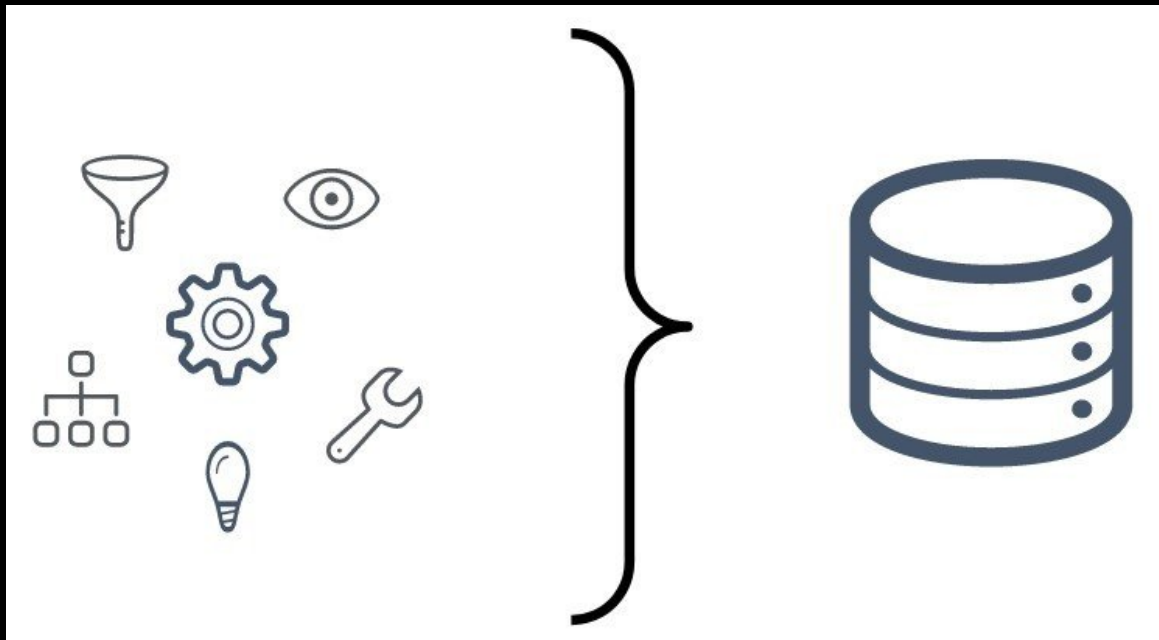


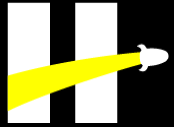
Imputación de Valores Faltantes



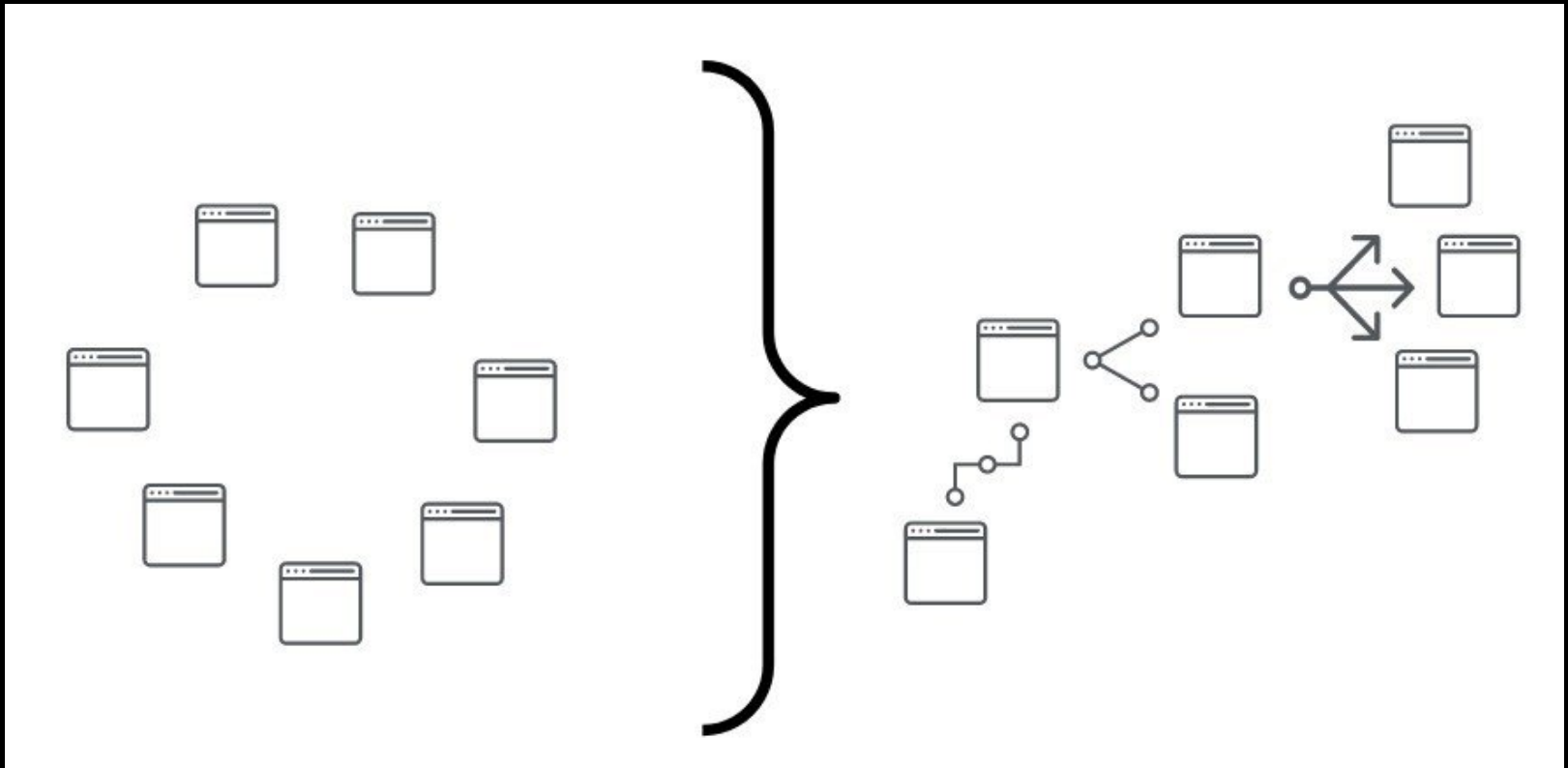


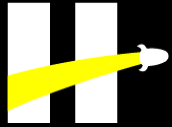
Transformación de Datos



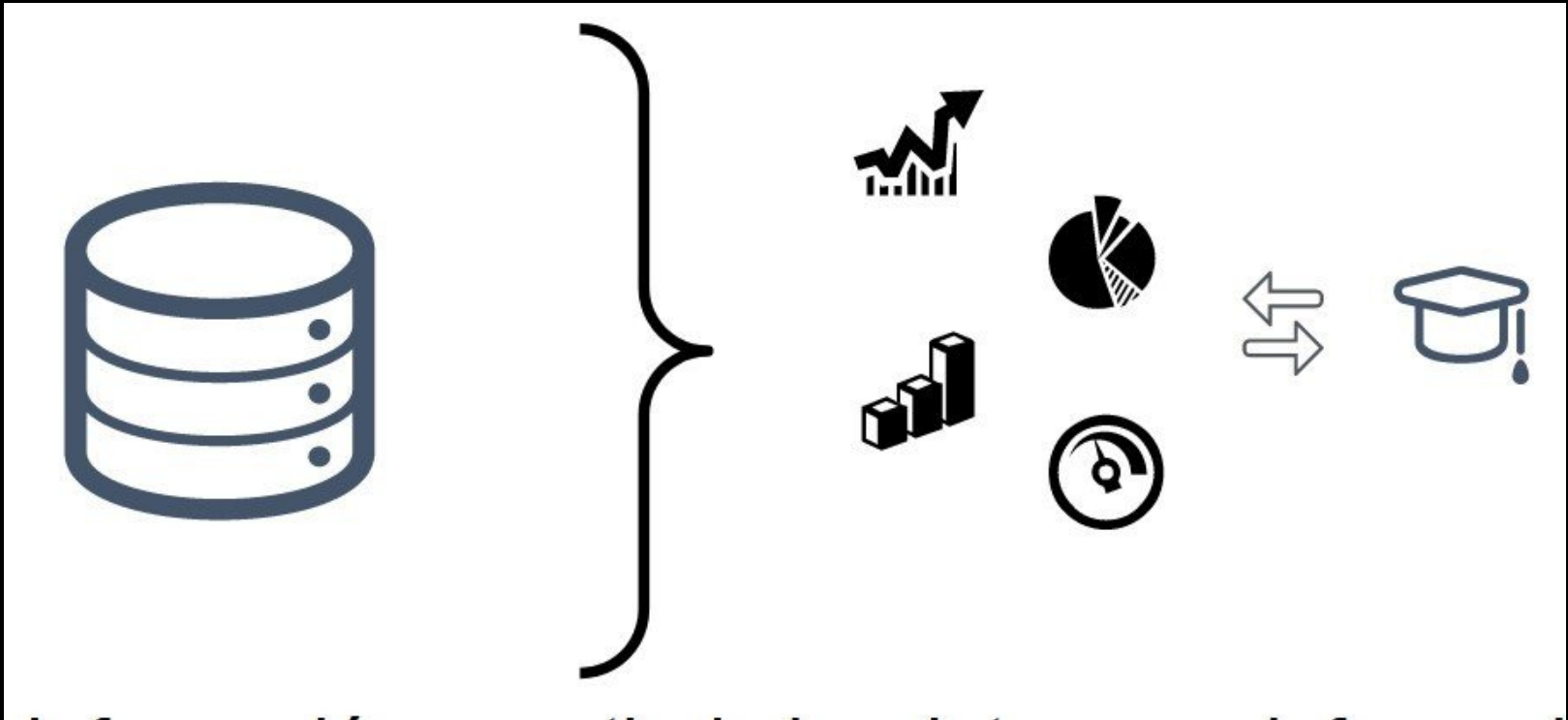


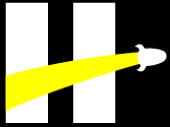
Modelado de Datos





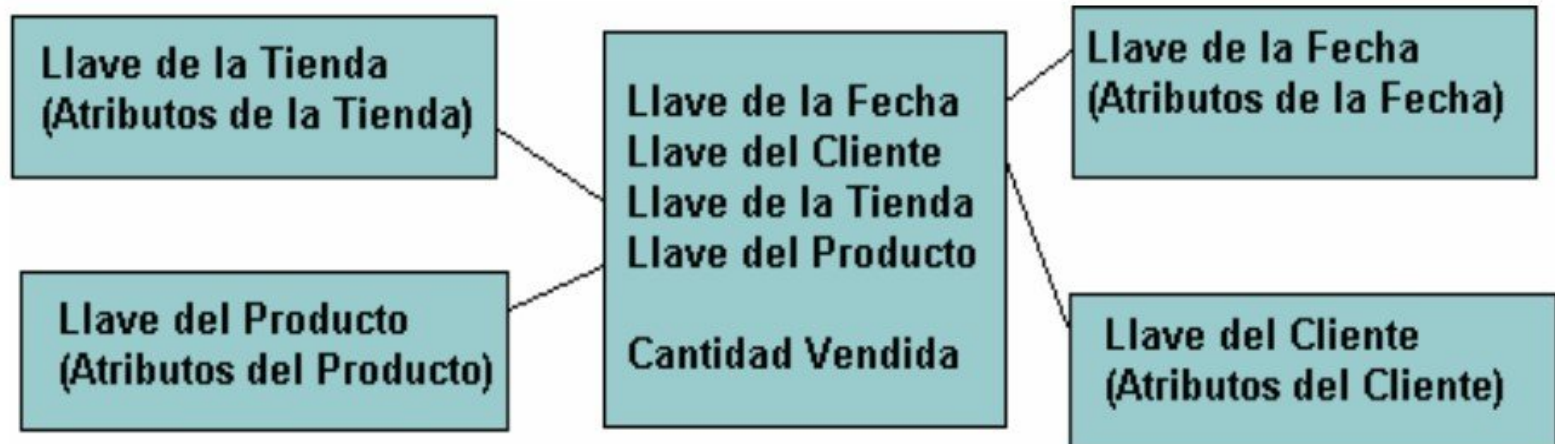
Reportes y Visualización

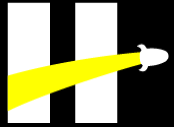




Hechos

Dimensiones





Claves Subrogadas

Una clave subrogada es un identificador único que se asigna a cada registro de una tabla. Puede obtenerse a partir de la conjunción de columnas ya preexistentes.