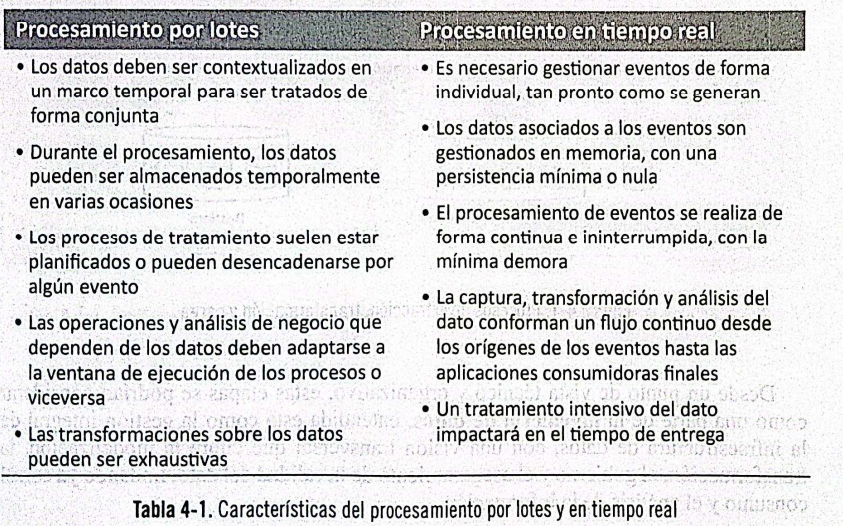
**TEMA 4**



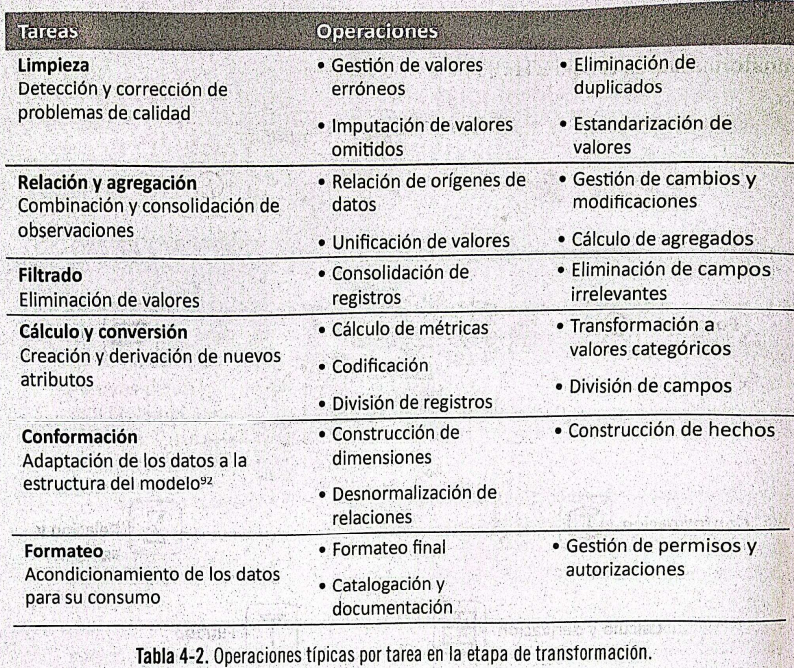
**ETL:** Extraccion Transformacion y Carga (Extract Load Transform)

**EXTRACCION:**

**Flujo ETL**: Consideraciones a la hora de implementar en esta etapa (**extraccion**)

* **Variedad de orígenes:** Pueden ser heterogéneas tanto en:
  + **Ubicación:** Locales o en la nube
  + **Formatdo:** Estructurado, semiestructurado, no estucturado y comprimido
  + **Tecnología:** Bases de datos propietarias, ERP, CRM, API, código abierto
  + **Propiedad:** Interna y externa
* **Extracciones totales y diferenciales:** Se usan 2 estrategias:
  + **Extraccion total:** Lee todo el volumen de datos en cada momento
  + **Extraccion diferenciales:**  Solo toma en cuenta los datos generados desde la última extracción. Implementa tecnología de captura de cambios (CDC)
* **Inventariado y perfilado:** Fundamental realizar un **perfil de los datos**. Consiste en un analisis de datos, permitiendo obtener estadísticas, valores anómalos, inconsistencias etc.

**TRANSFORMACIÓN**:



**Microservicios**: Ensamblado de operaciones ejecutadas en sistemas dispersos, gestionados en la nube (*serverless*)

**CARGA:**

**Replicación de datos:** Permite tener en sicronia estructuras de datos en distintos repositorios. Tiene impacto mínimo en sistemas origen ya que el control de cambios se realiza en el log

**CDC**: Se utiliza fundamentalmente en la capa de extracción como mecanismo de detección de cambios y acceder a ellos para transformarlos