



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

IIC3103

Taller de Integración

Profesor

Arturo Tagle / Daniel Darritchon



Resumen Clase Anterior

Integración por datos

1. Transferencia de archivos
2. RPA: una solución a este problema
3. Caso de estudio BancoEstado



Arquitecturas de integración

Módulo 4 - Clase 7

Integración por datos

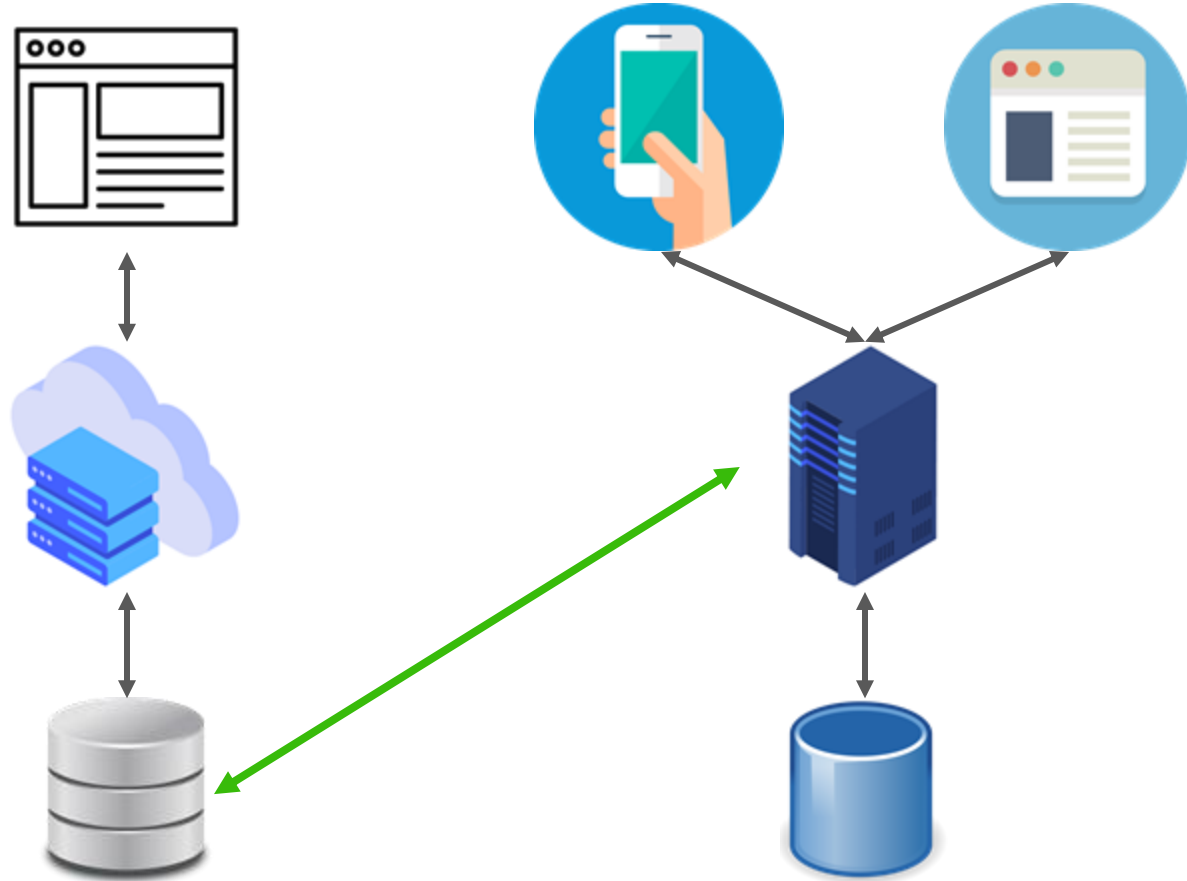
Parte 2

2. Lectura cruzada de fuentes de datos

Leyendo datos de otros sistemas

Arquitecturas de integración de datos

A nivel de capa de datos



Arquitecturas de integración de datos

A nivel de capa de datos

Ventajas

- Acceso en tiempo real a la información.
- Eliminación de redundancia de datos.
- Eficiencia de la comunicación.

Desventajas

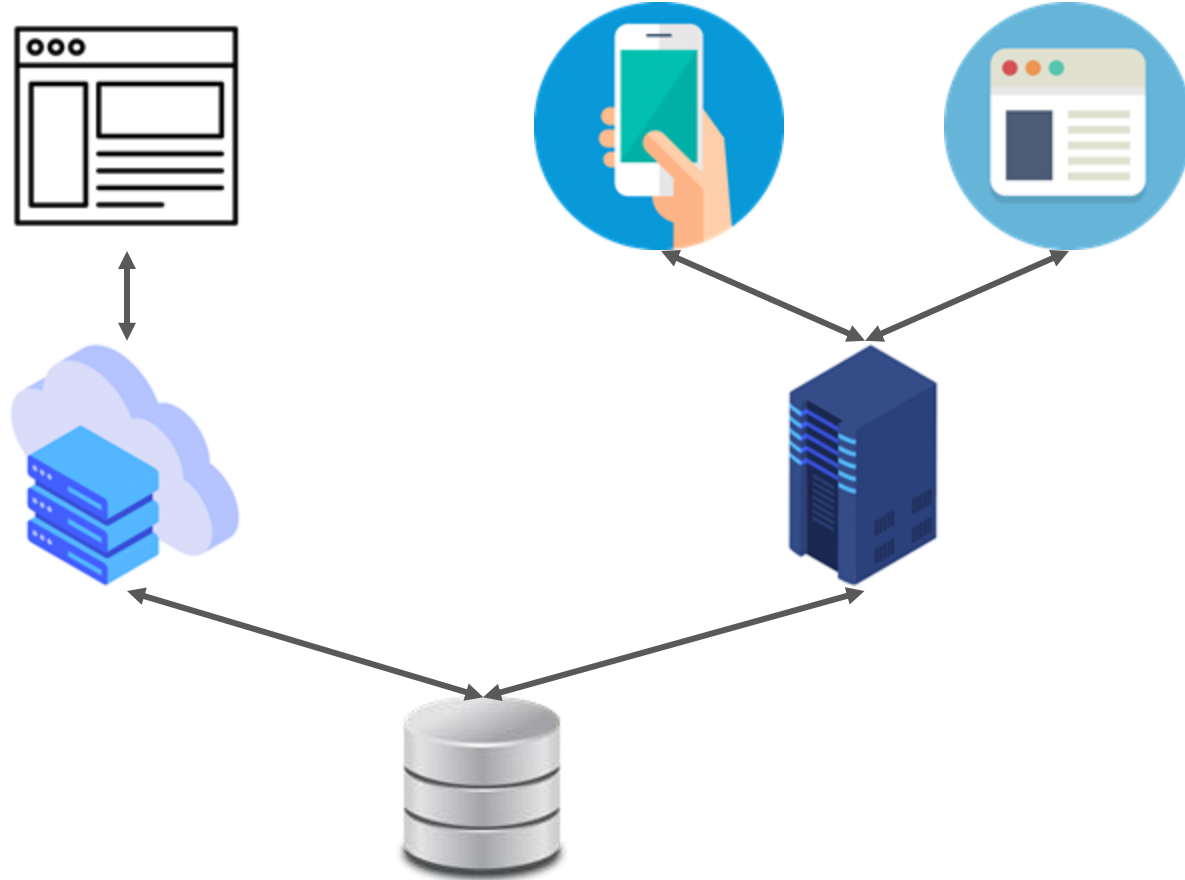
- Seguridad de los datos.
- Problemas de integridad de datos.
- Dificultades de mantenimiento.
- Riesgo de sobrecarga.

3. Base de datos compartida

Unificando la base de datos

Arquitecturas de integración de datos

A nivel de capa de datos



Arquitecturas de integración de datos

A nivel de capa de datos

Ventajas

1. Consistencia de datos.
2. Economía de recursos.
3. Coherencia en tiempo real.

Desventajas

1. Cuellos de botella y problemas de rendimiento.
2. Problemas de seguridad.
3. Dificultades de mantenimiento y escalabilidad.
4. Problemas de integridad de datos.

¿Qué tan escalable es este modelo?

Factores a considerar:

1. Carga de la base de datos. (R/W)
2. Consistencia de los datos.
3. Coordinación de cambios.
4. Aislamiento.

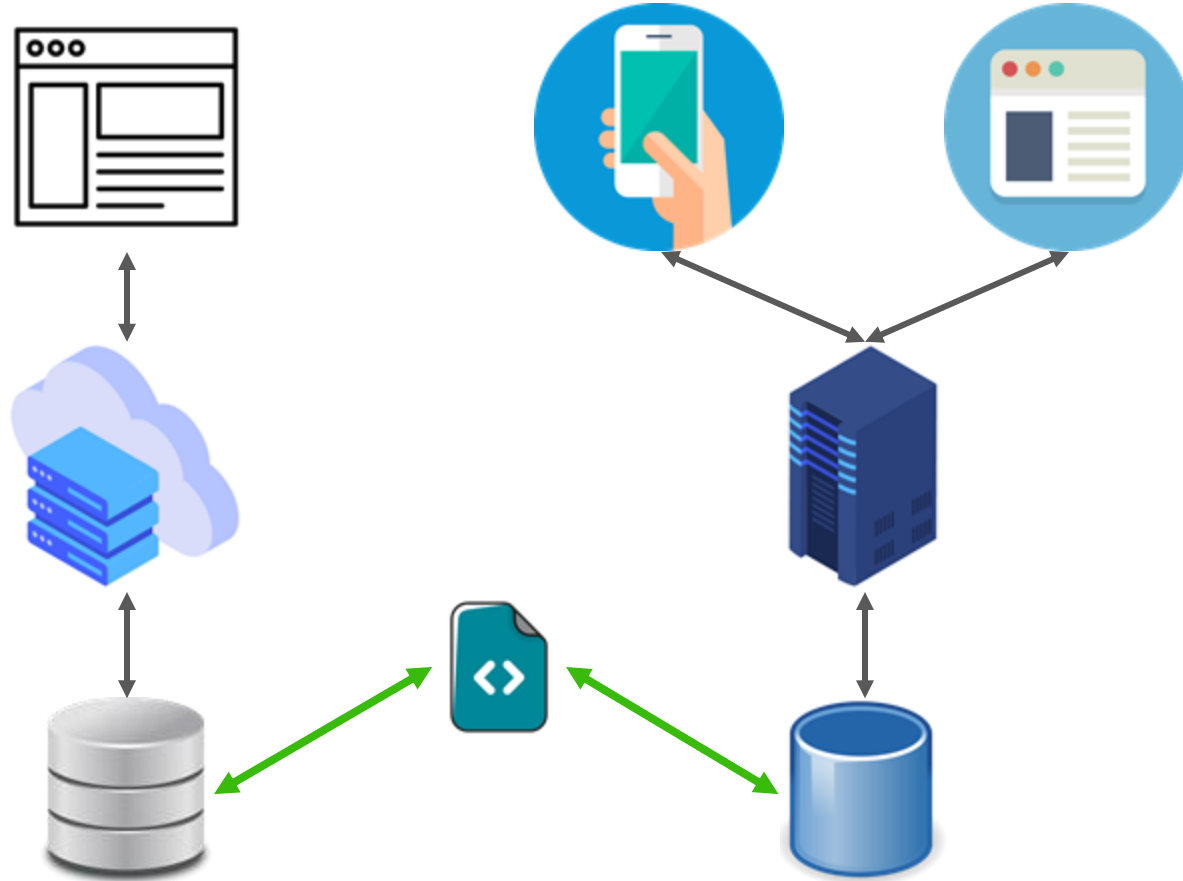


4. Procesos de datos

Extrayendo, transformado e insertando datos

Arquitecturas de integración de datos

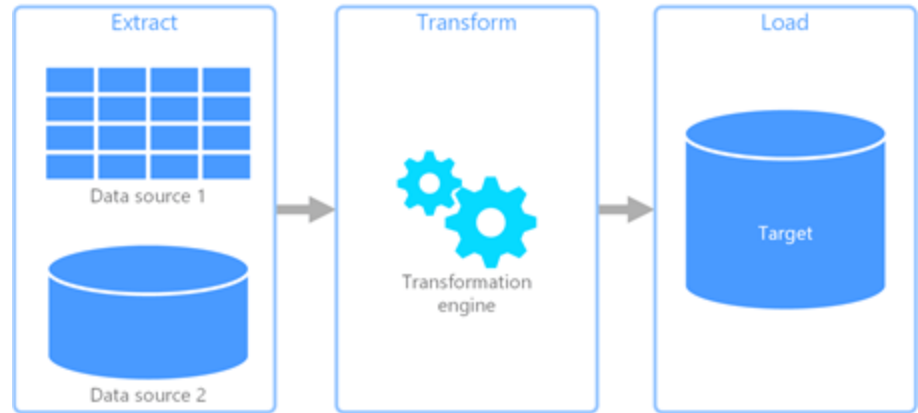
A nivel de capa de datos



Procesamiento de datos: ETL

Un ETL es un proceso de extracción, transformación y carga de datos a una nueva fuente de datos

Normalmente se relacionan con procesos de carga de información en Datawarehouses



Procesamiento de datos: ETL

Casos de usos comunes:

- Extraer información de sistemas transaccionales a bases de analítica
- Migración de sistemas
- Integración de datos entre sistemas



Ejemplo ETL: Tablas Agregadas

Tabla
Agregada

Store_ID	Item_ID	Month_ID	Revenue
1	10	199903	1305584

Query pulls from the
aggregate table

Month of Year	Metrics	Revenue
March		\$ 1,305,584

Pre-aggregation
creates the aggregate
table

Tabla Original

Store_ID	Item_ID	Day_ID	Revenue
1	10	3/9/99	16,152
1	10	3/10/99	28,541
1	10	3/11/99	80,892
1	10	3/12/99	51,200
...

Cuando arquitecturas
de datos se encuentran
con arquitecturas
basadas en eventos

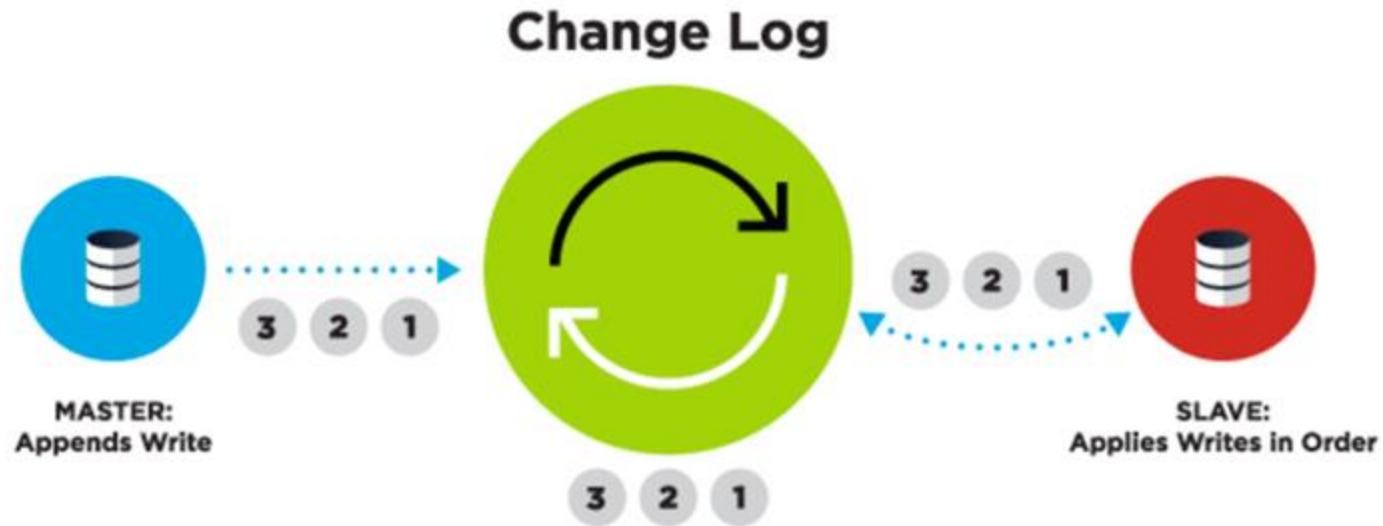


What is Change Data Capture?

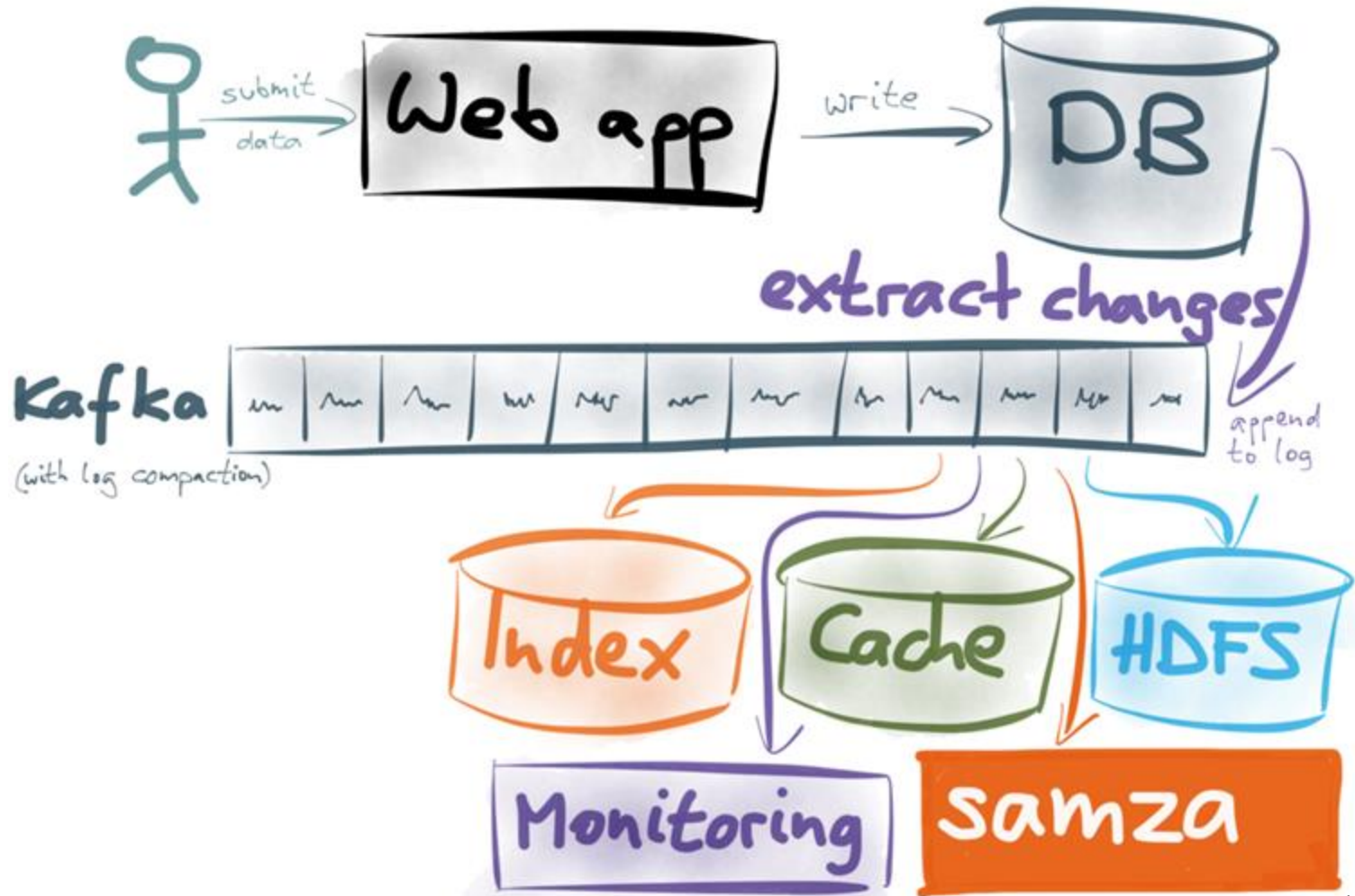


What is Change Data Capture?

Change Data Capture



USING CHANGE CAPTURE

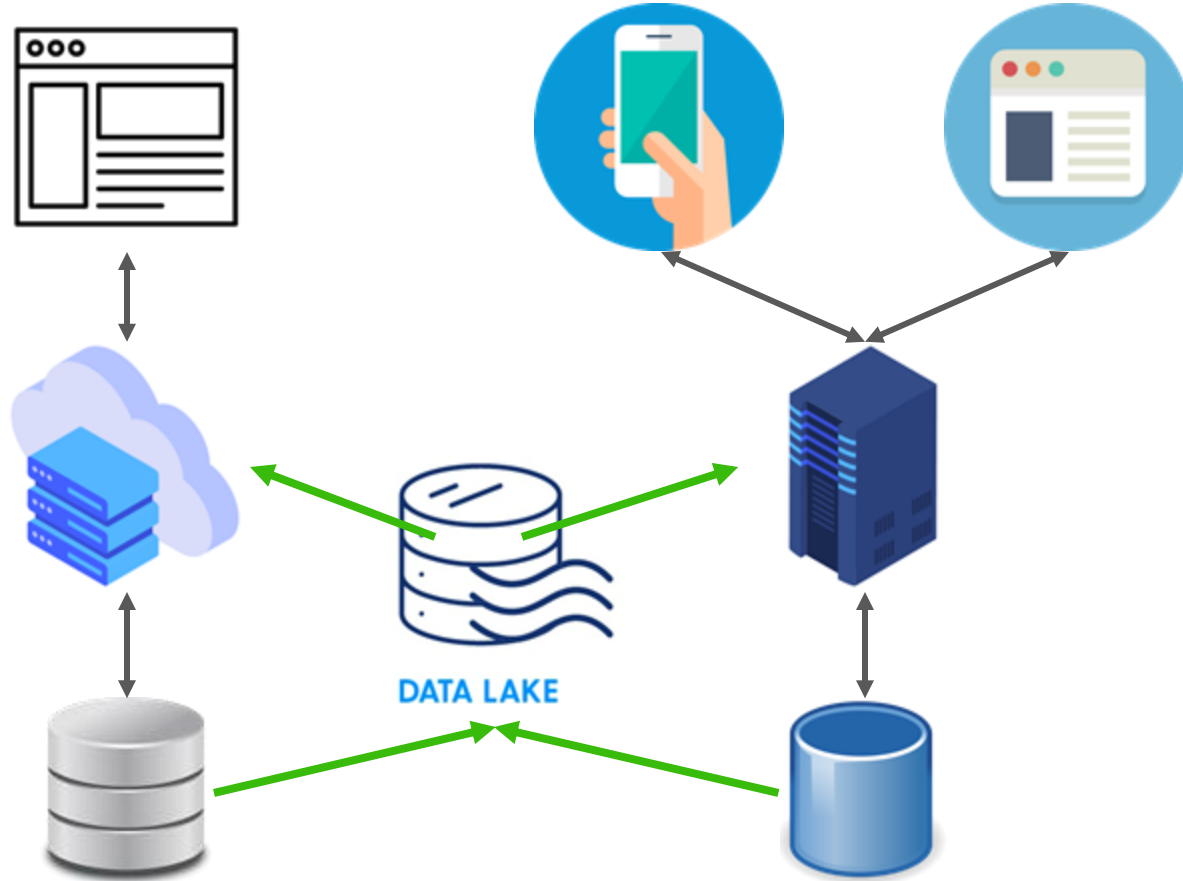


5. Datos centralizados

Fuente única de información

Arquitecturas de integración de datos

A nivel de capa de datos



Why the Market is Shifting from ETL to EL→T



ETL vs ELT

Qlik®
LEAD WITH DATA

Tecnologías de Datalake han impulsado el cambio desde ETL a ELT

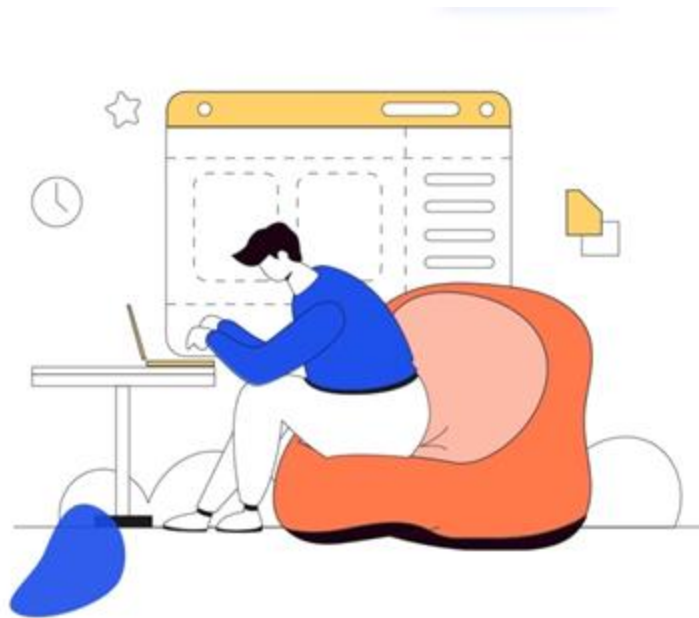
Almacenamiento de datos sin procesar y dejar al consumidor final de los datos la transformación de datos.

El Datalake es sólo el almacenamiento de la información. Para poblar, se requieren procesos de captura de información.

Existen arquitecturas escaladas de Datalake como data ocean o data mesh.



Actividad 7





PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

IIC3103

Taller de Integración



Bibliografía

Desafíos y formas de integración

Hohpe, G. (2002, July). Enterprise integration patterns. In 9th conference on pattern language of programs (pp. 1-9).

Hohpe, G., Woolf, B. (2015). Introduction. In *Enterprise integration patterns: Designing, building and deploying messaging solutions* (pp. xxix - li). Boston, Massachusetts: Addison-Wesley.

Fowler, M. (2015). Patterns of enterprise application architecture. Addison-Wesley.

ETL y CDC

Mohanan, S. (2021, March 14). *Change data capture (CDC) for ETL: 3 easy steps*. Hevo Data. Retrieved April 25, 2022, from <https://hevo.com/learn/change-data-capture/>