

Proyectos de grado 2024

Grupo GEMA

Gestión, Modelado y Análisis de datos

Arquitecturas para el análisis de Big Data

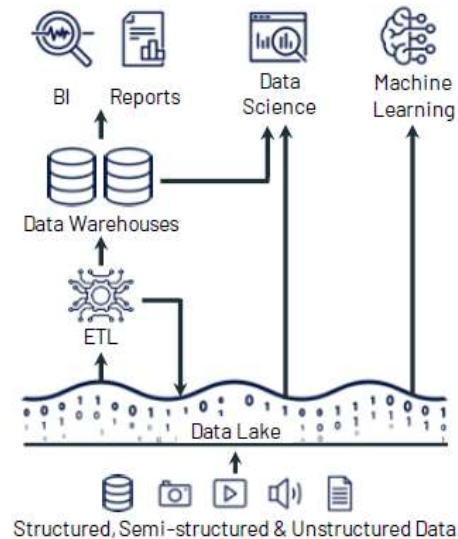
Lakehouse: A New Generation of Open Platforms that Unify DataWarehousing and Advanced Analytics

CIDR '21, Jan. 2021, Online

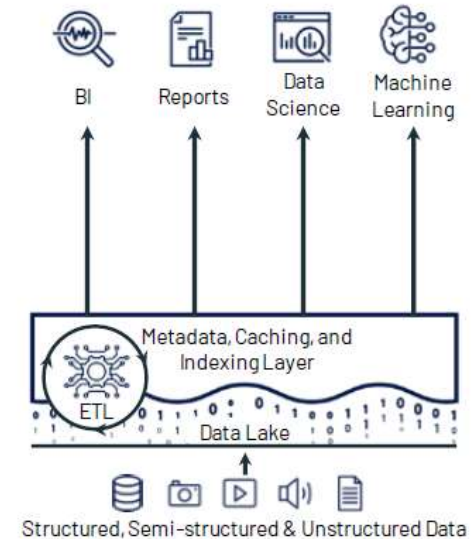
Michael Armbrust, Ali Ghodsi, Reynold Xin, and Matei Zaharia



(a) First-generation platforms.

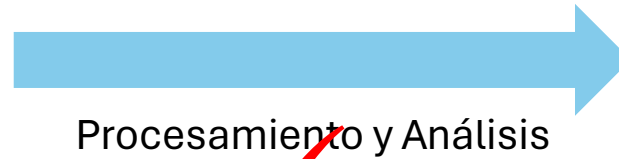
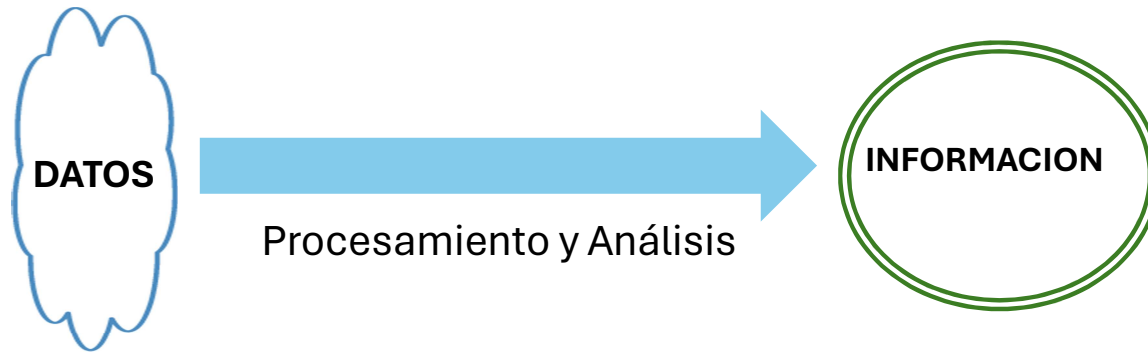


(b) Current two-tier architectures.



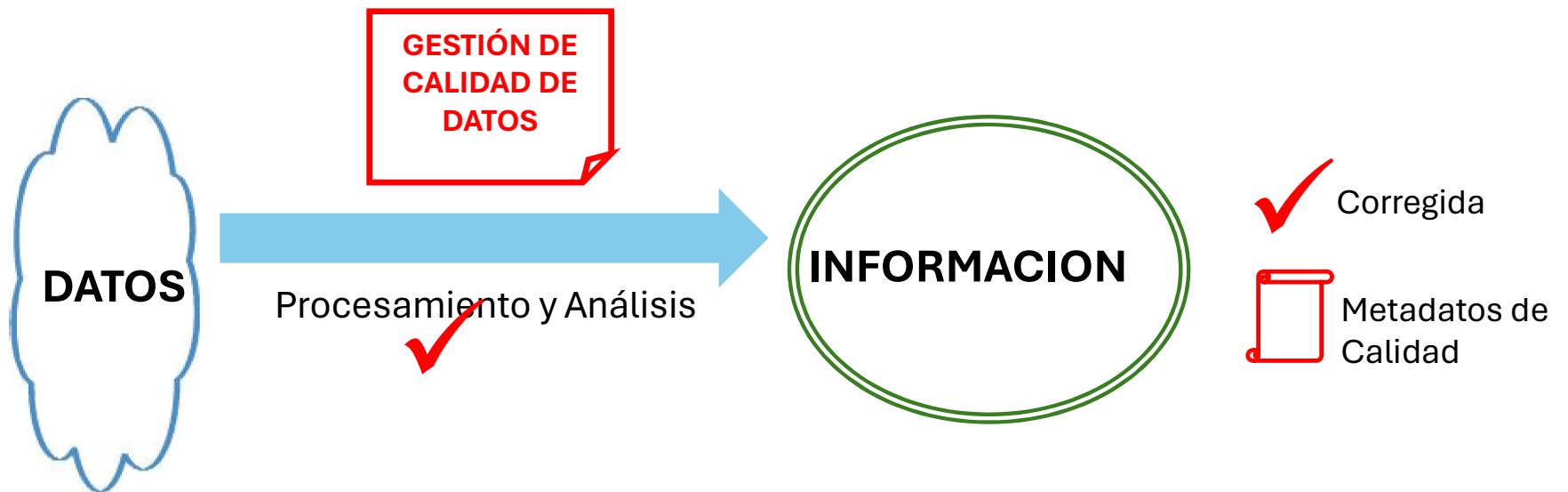
(c) Lakehouse platforms.

3 Calidad de Datos



Garbage In –
Garbage Out
phenomenon

4 Calidad de Datos

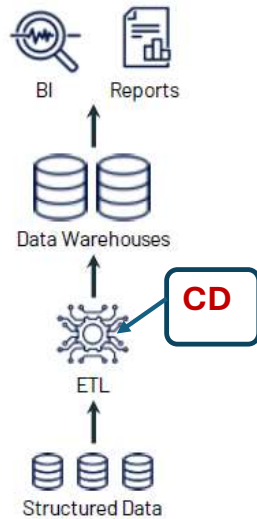


Calidad de Datos en Big Data: Desafíos

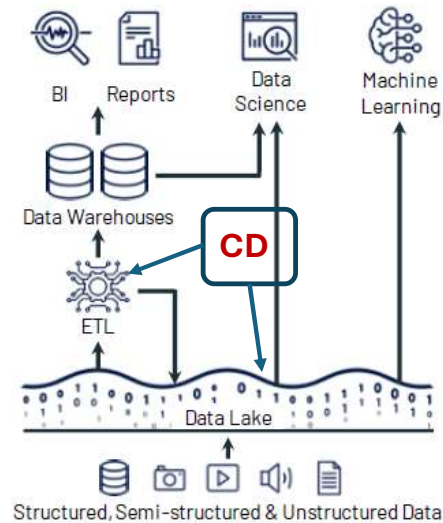
- Arquitecturas de Big Data
 - *Data Preparation* para Análisis de Big Data
- Modelos de datos y formatos heterogéneos
 - Estructurados, semi-estructurados, no estructurados
 - Granularidades de datos. Difíciles de definir debido a la **falta de esquema en los datos.**
- **Dimensiones de calidad**
 - Cuáles son más relevantes en este contexto?
 - Cómo se definen y cómo se miden?

Arquitecturas de Big Data

CIDR '21, Jan. 2021, Online

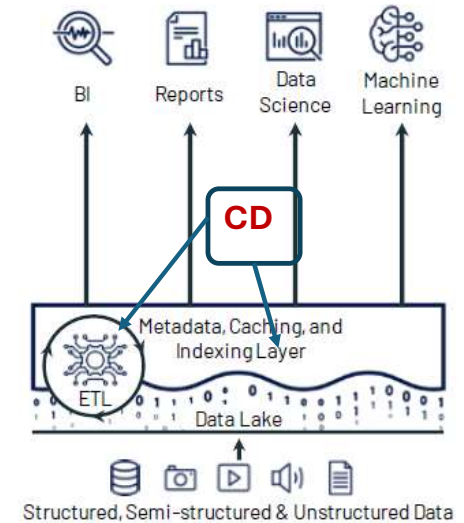


(a) First-generation platforms.



(b) Current two-tier architectures.

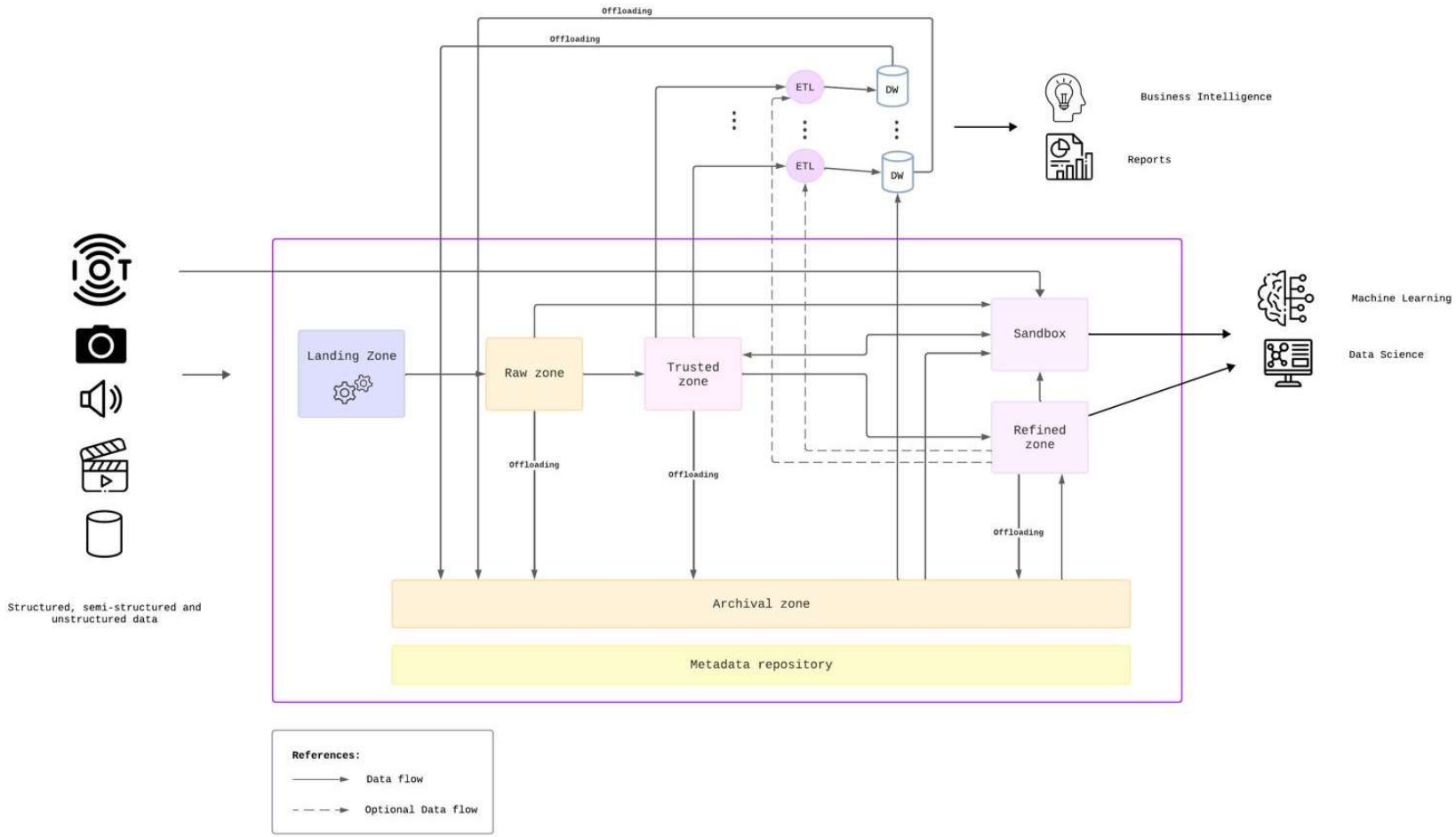
Michael Armbrust, Ali Ghodsi, Reynold Xin, and Matei Zaharia



(c) Lakehouse platforms.

Data Preparation

Trabajo en curso



Proyectos de grado 2024

- Herramienta para la aplicación de metodologías de gestión de Calidad de Datos
- Evaluación de exactitud sobre Data Lakes considerando el contexto
- Hacia un modelo de datos global en Data Lakes: caso datos abiertos de gobierno
- Calidad de datos en bases de datos de grafos