

Data Lakes 101

Carolina Acosta Tovany

#Devday4w



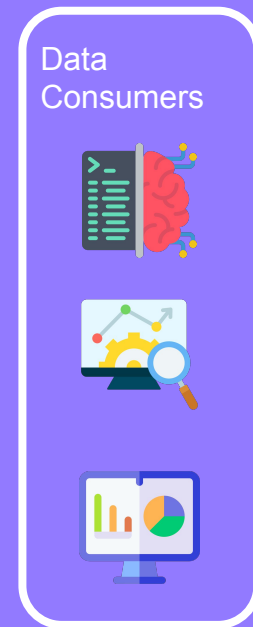
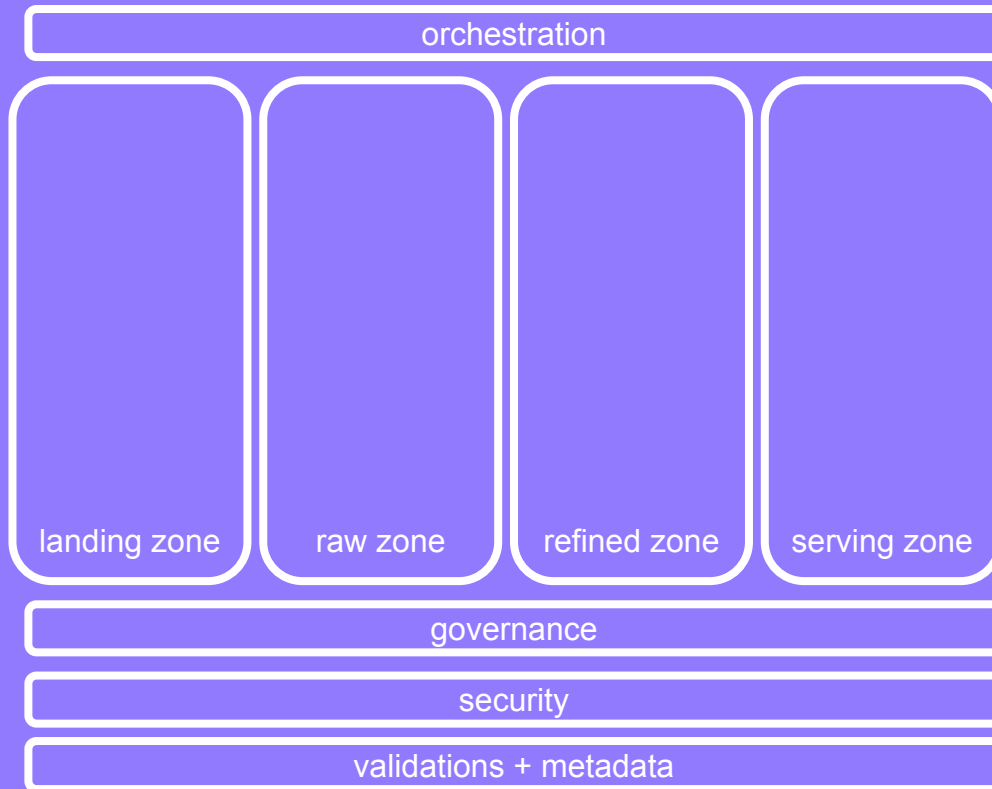
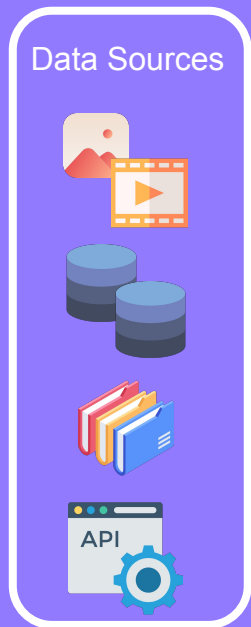
¿Data Lake?



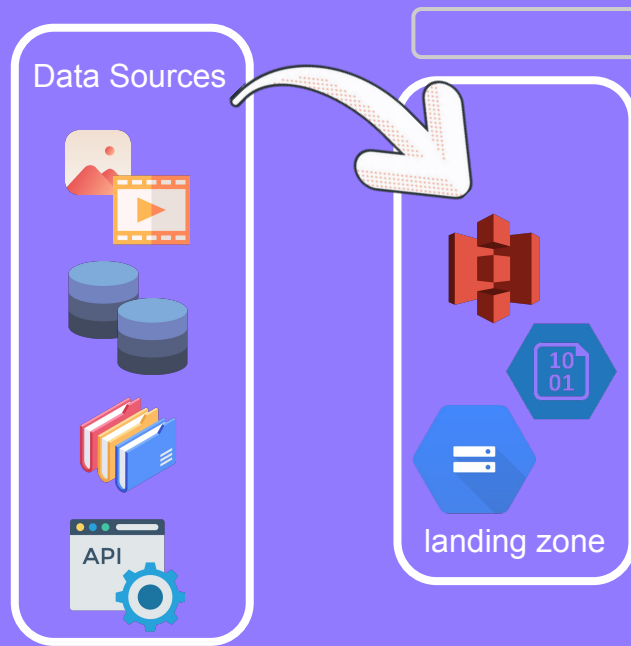
Ventajas



Arquitectura básica de un Data Lake



Arquitectura básica de un Data Lake



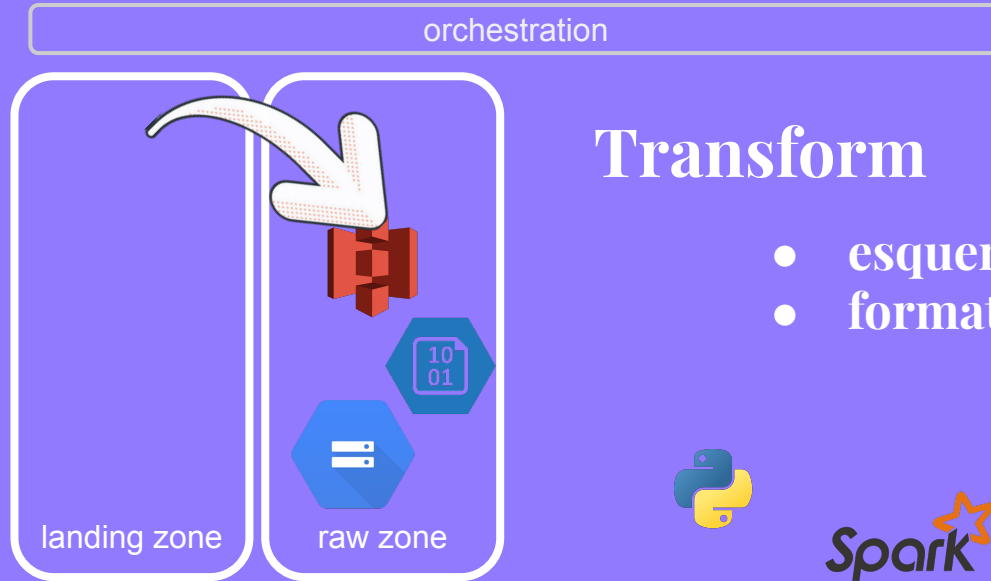
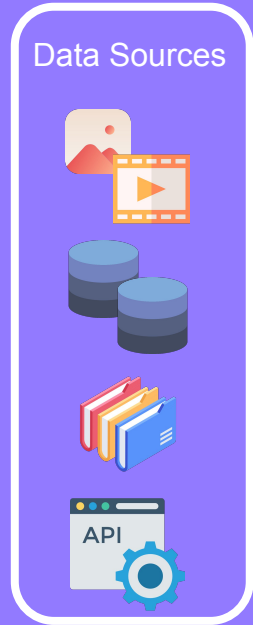
orchestration



Ingesta

- ETL (Extract Transform Load)
- ELT (Extract Load Transform)

Arquitectura básica de un Data Lake

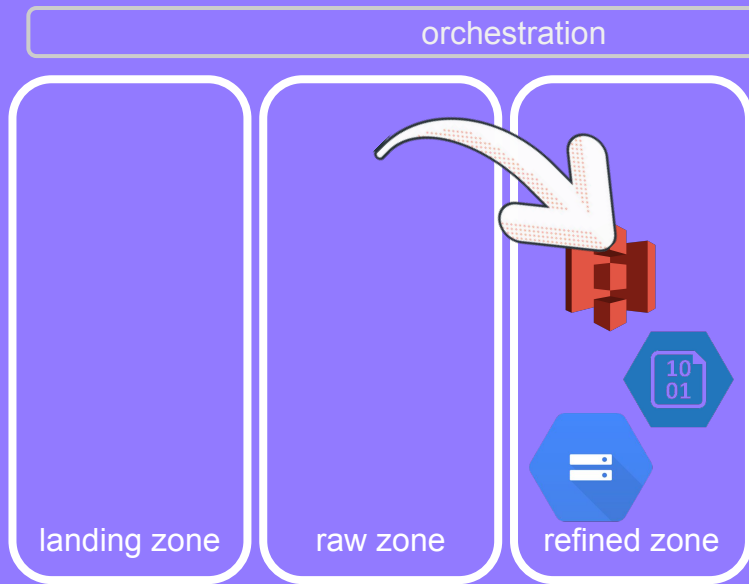
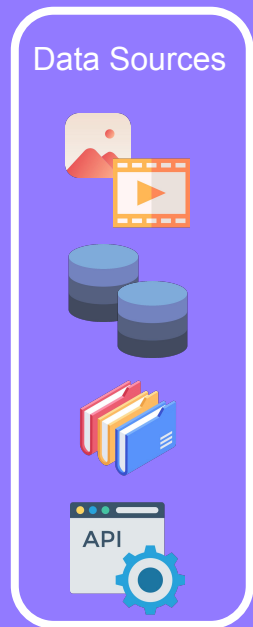


Transform

- esquema
- formato



Arquitectura básica de un Data Lake

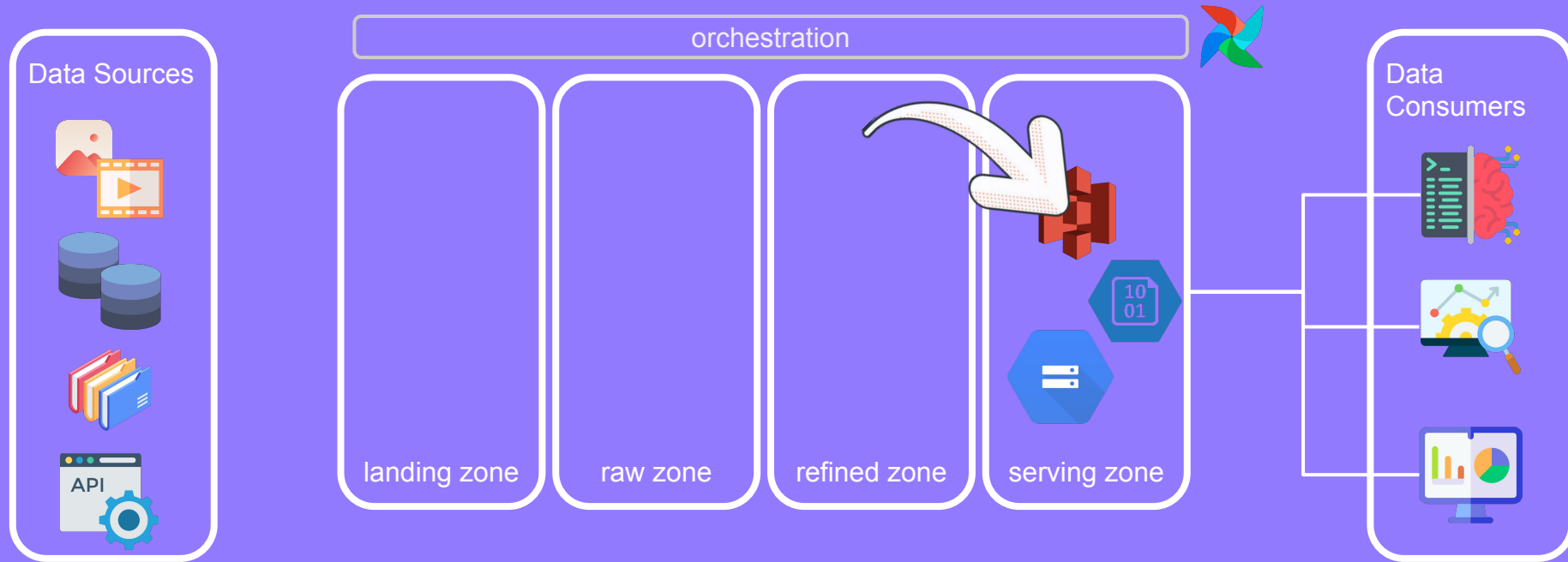


Transform

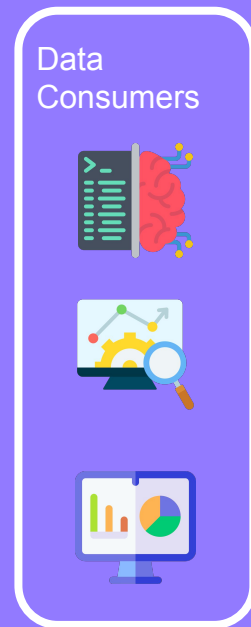
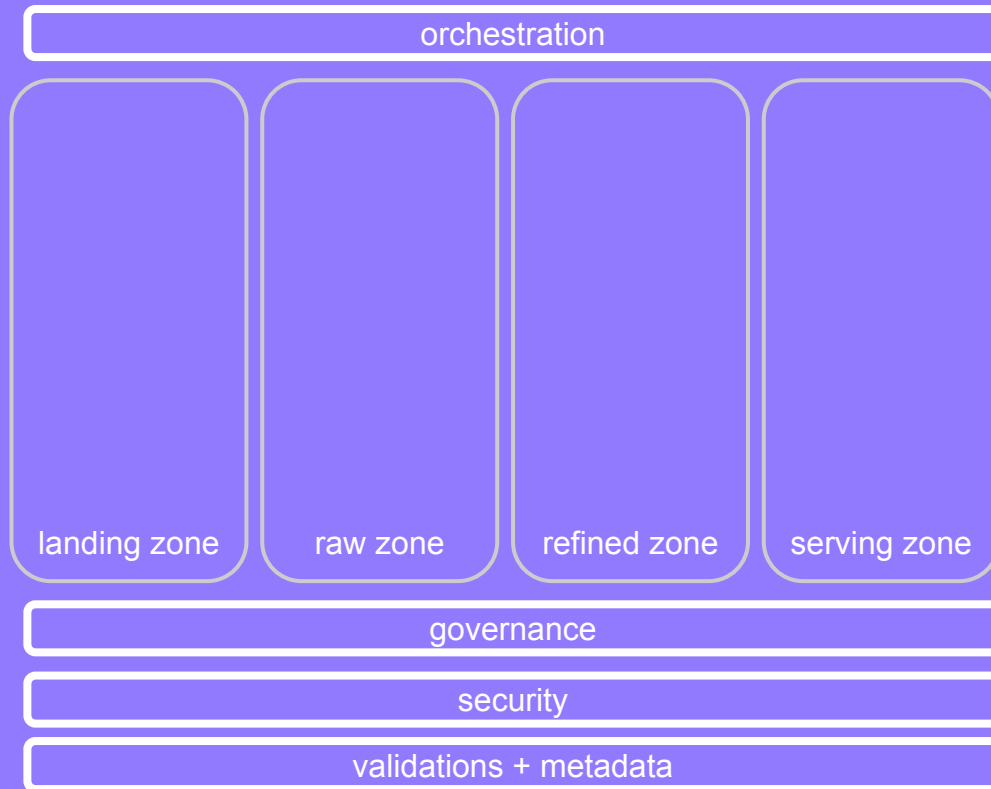
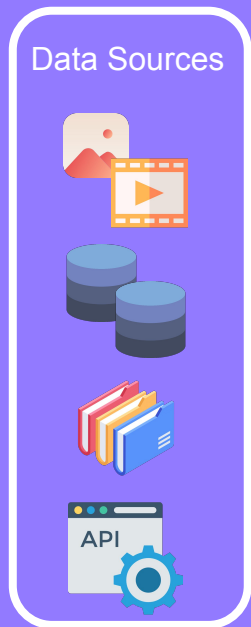
- modelos ML
- reglas de negocio



Arquitectura básica de un Data Lake



Arquitectura básica de un Data Lake



Ejemplo

Data source





Landing zone

cumulo=18/
rundate=2022-06-10



landing zone





Raw zone

cumulo=18/
rundate=2022-06-10



raw zone





Refined zone

cumulo=18/
rundate=2022-06-10/
fauna=true/



*.jpg

fauna=false/
*.jpg

refined zone





Serving zone

cumulo=18/



fauna=true/
*.jpg

fauna=false/
*.jpg

serving zone



Conclusiones

- Repositorio de datos versátil
- Es escalable — Big data friendly
- Tiene diferentes capas (de acuerdo al caso de uso)
- Diseñado para muchos perfiles de consumo

¡Gracias!



@caroacostatovany

in

