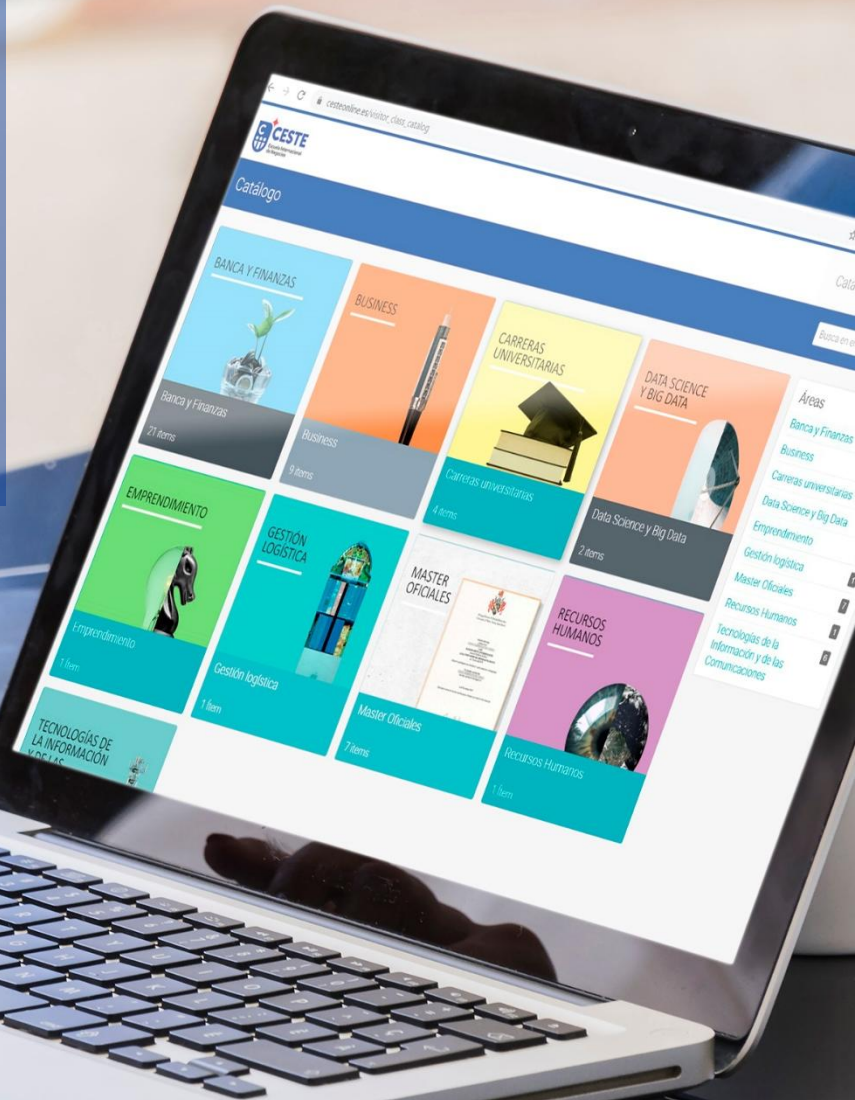


CESTE

Escuela Internacional de Negocios

Zaragoza (España)



3. Diseño de pipelines escalables y sostenibles en Databricks

Índice | Diseño de pipelines escalables y sostenibles en Databricks

1. Buenas prácticas en diseño de pipelines
2. Orquestación con Databricks Workflows
3. Control de costes y clusters eficientes
4. Casos reales

Objetivos | Pipelines modulares, eficientes y de bajo coste

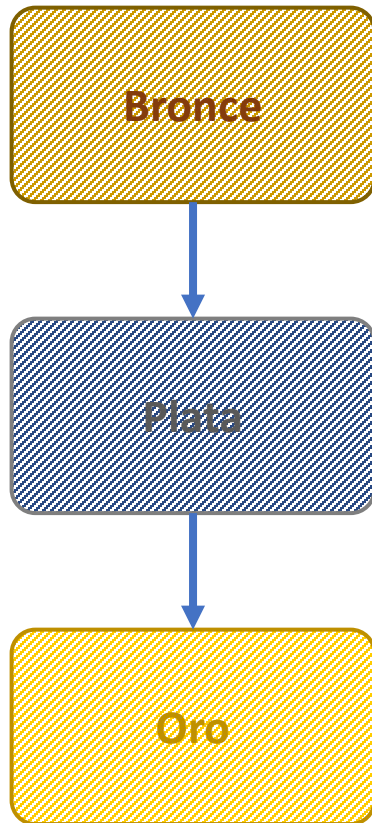


1. Diseñar pipelines modulares y resilientes en Databricks siguiendo buenas prácticas
2. Optimizar costes y recursos en producción, aprendiendo con casos reales de optimización.
3. Orquestar y automatizar pipelines con Workflows, integrando parámetros, dependencias y control de versiones

Buenas prácticas | Modular, parametrizable y resiliente a errores

Capa	Staging / Bronce		Core / Silver	Presentation / Gold	
Sublayer	Raw	Standard	Conformed	Datamart	Report
Propósito	<ul style="list-style-type: none">• Recogida de las fuentes de datos• Forma original• Auditoría y recuperación de errores	<ul style="list-style-type: none">• Datos limpios• Preparación inicial para análisis	<ul style="list-style-type: none">• Entidades de negocio (cliente, producto, transacción, ...)• Datos enriquecidos• Agregaciones o cálculos	<ul style="list-style-type: none">• Diseñada para ML• Variables listas para entrenamiento	<ul style="list-style-type: none">• Listo para aplicaciones finales• Necesidades de negocio
Consumidor	Data Engineering	Data analyst, Engineers	Data Engineers, analyst Scientist	Data Scientist, BI	"C" Roles

Buenas prácticas | Modular, parametrizable y resiliente a errores

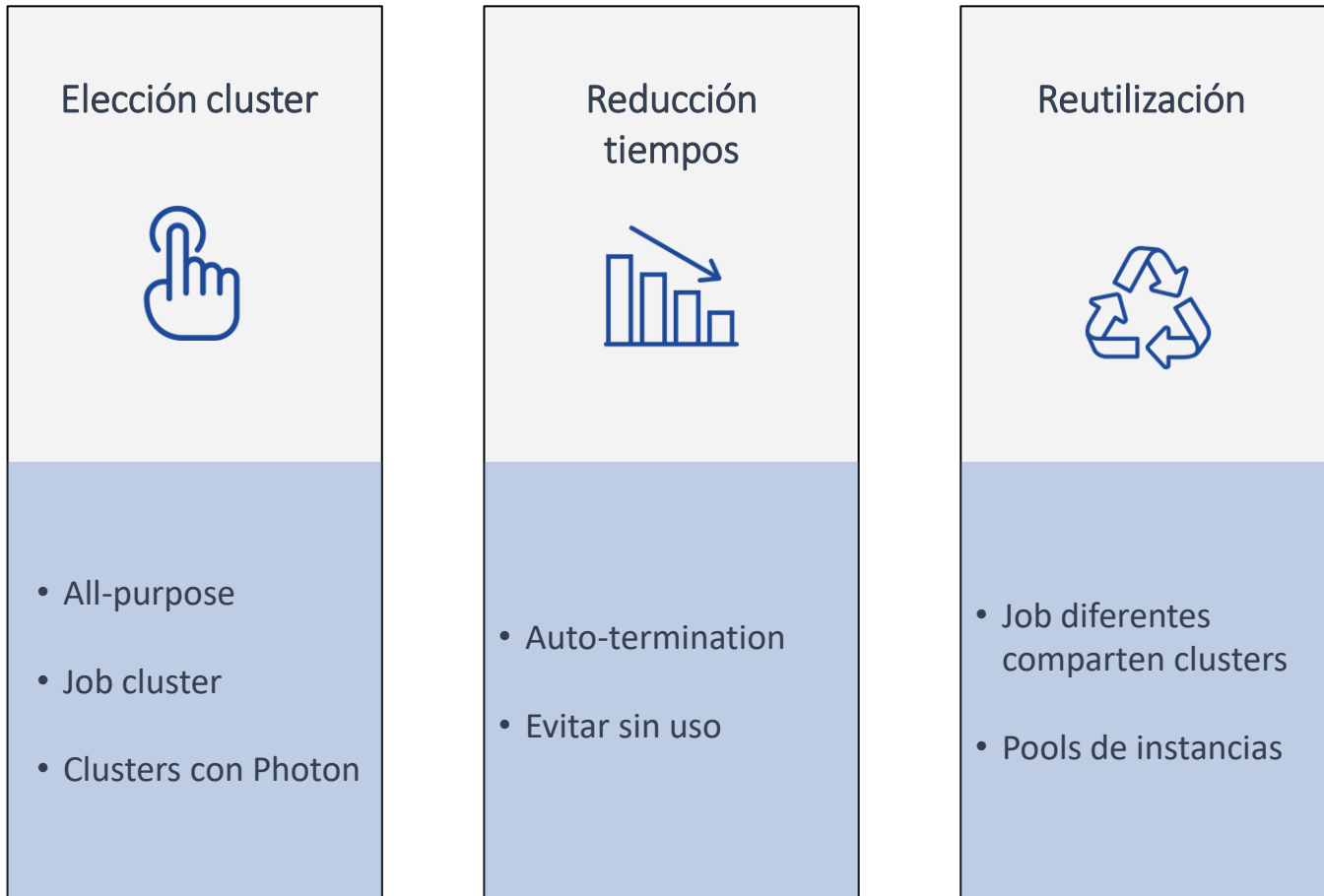


- Evitar hardcodeos
- Usar ficheros JSON/YAML



- Manejo try/except
- Reintentos configurables en Databricks
- Alertas

Control de costes | Un pipeline sostenible no solo se optimiza en rendimiento, también en coste de ejecución.



Orquestación en Databricks | Orquestar es más que ejecutar: es
conectar tareas, parametrizarlas y mantenerlas versionadas



databricks



www.ceste.es