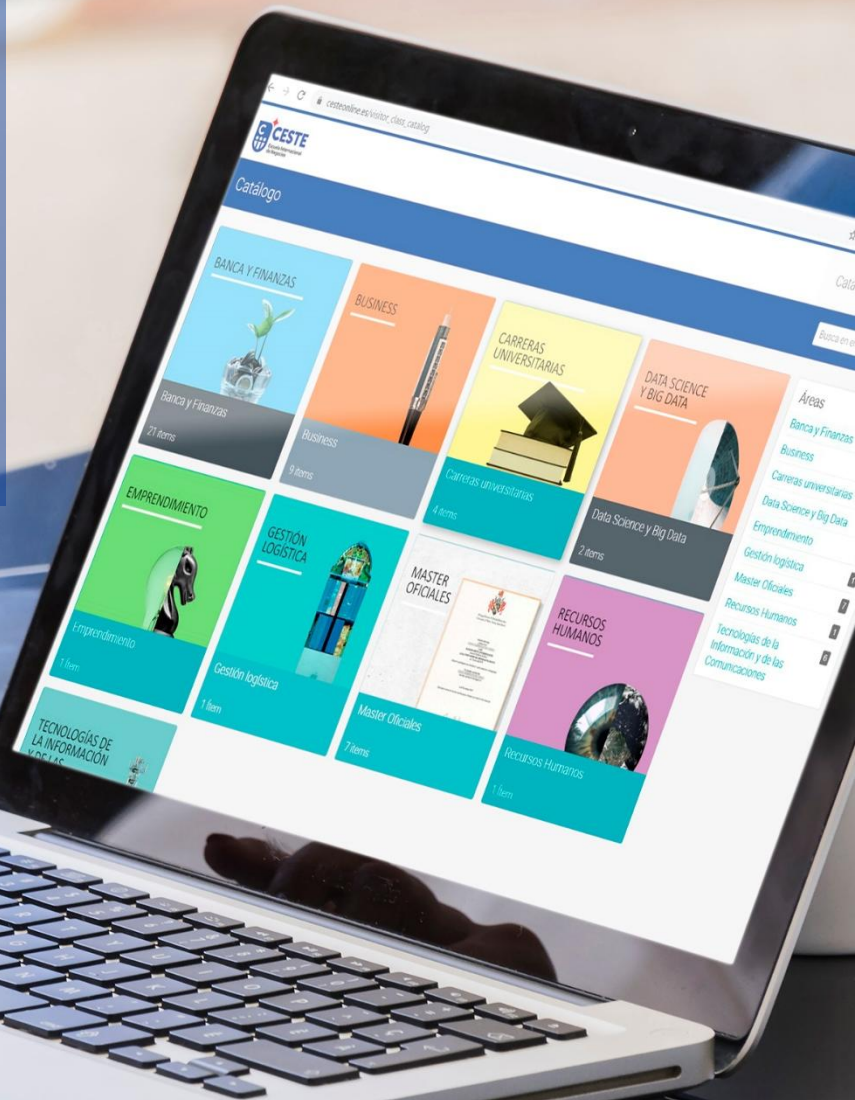


CESTE

Escuela Internacional de Negocios

Zaragoza (España)



3. Databricks en entornos productivos: proyecto E2E

Índice | Databricks en entornos productivos

1. Arquitectura de capas
2. Modelado de datos
3. Proyecto E2E

Objetivos | Construir ETLs basadas en Delta Lakes

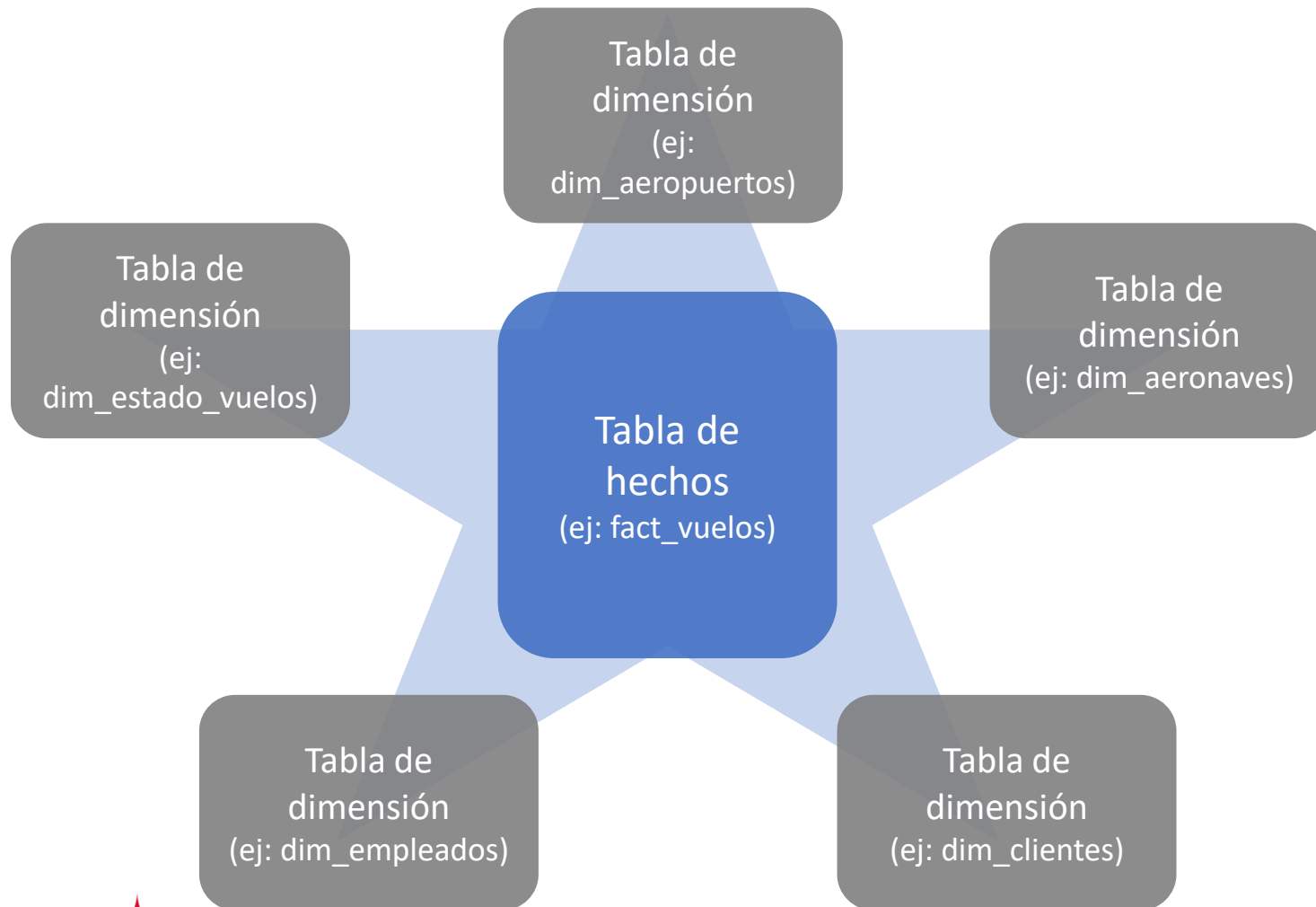


1. Diseñar un pipeline de ingesta y transformación de datos utilizando el enfoque medallion
2. Configurar y ejecutar un Workflow con múltiples tareas dependientes usando Databricks Jobs
3. Aplicar principios de buenas prácticas y gobierno de datos

Arquitectura de capas | Del dato crudo al valor con eficiencia

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

Modelado de datos | Patron de estella para relacionarlos



Buenas prácticas en Producción | Puntos a tener en cuenta para mantener en un entorno en perfecto estado

Separación de entornos



- DEV / STAGGING / PROD
- Mediante Workspace separados

Configuraciones externas



- Ficheros configs
- Variables entorno o secret

Uso eficiente de clústeres



- Job clusters
- Autoscaling

Manejo de errores



- Reintentos automáticos
- Logging

Versionado



- Repos

Proyecto E2E completo | Como construir un proyecto E2E en un ambiente productivo



databricks



www.ceste.es