



CESTE

Escuela Internacional de Negocios

Zaragoza (España)

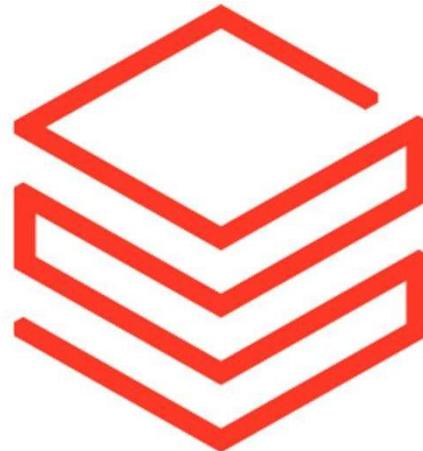
The screenshot shows the CESTE online catalog interface. The main header reads "Catalogo". Below it, there are several colored boxes representing different academic areas:

- BANCA Y FINANZAS: Banca y Finanzas (21 items), Business (9 items)
- BUSINESS: Carreras universitarias (4 items)
- CARRERAS UNIVERSITARIAS: Data Science y Big Data (2 items)
- DATA SCIENCE Y BIG DATA: Emprendimiento (1 item), Gestión logística (1 item), Master Oficiales (7 items), Recursos Humanos (1 item)
- EMPRENDIMIENTO: Gestión Logística (1 item)
- GESTIÓN LOGÍSTICA: Master Oficiales (7 items)
- MASTER OFICIALES: Recursos Humanos (1 item)
- RECURSOS HUMANOS: TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES (1 item)

On the right side, a sidebar titled "Áreas" lists the categories: Banca y Finanzas, Business, Carreras universitarias, Data Science y Big Data, Emprendimiento, Gestión logística, Master Oficiales, Recursos Humanos, and Tecnologías de la Información y las Comunicaciones.

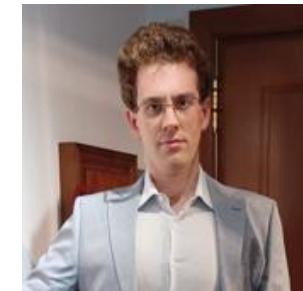


Introducción a Databricks y su Arquitectura de Clusters



databricks

¿Quién soy?



Rodrigo Anchelergues Cortina

Graduado en Estudios en Arquitectura (Universidad de Zaragoza)
Master en Ciencia de Datos e Inteligencia Artificial (CESTE)

Profesional más de 3 años de experiencia entorno al mundo de los datos.

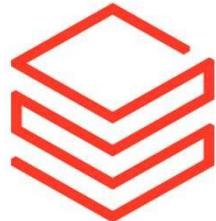


Objetivos de la sesión



- Breve historia de Databricks
- Entender Databricks y su arquitectura.
- Identificar los tipos de clusters y su uso.
- Familiarizarse con el entorno de trabajo.
- Creación de un cluster y navegar por el entorno.

Historia de Databricks



databricks

- Fundada por creadores de Apache Spark (2013)
- Surgió del Proyecto AMPLab (<https://amplab.cs.berkeley.edu/>)
- Fundadores: Ali Ghodsi , Andy Konwinski.
- En 2017 fue anunciada como servicio de Azure ([Azure Databricks](#))

Actualidad Plataformas de Datos



ETLs



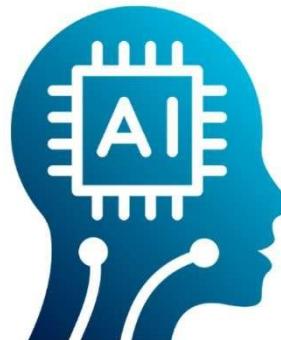
Data
Warehouse



Business
Intelligence



Plataforma

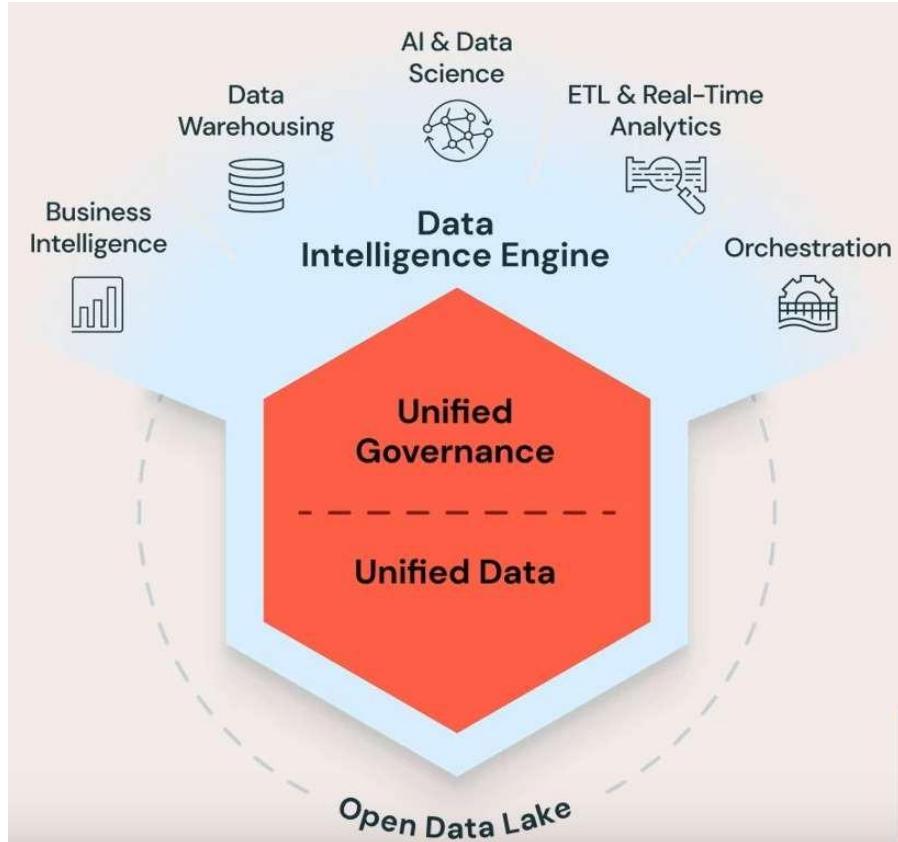


Inteligencia
Artificial



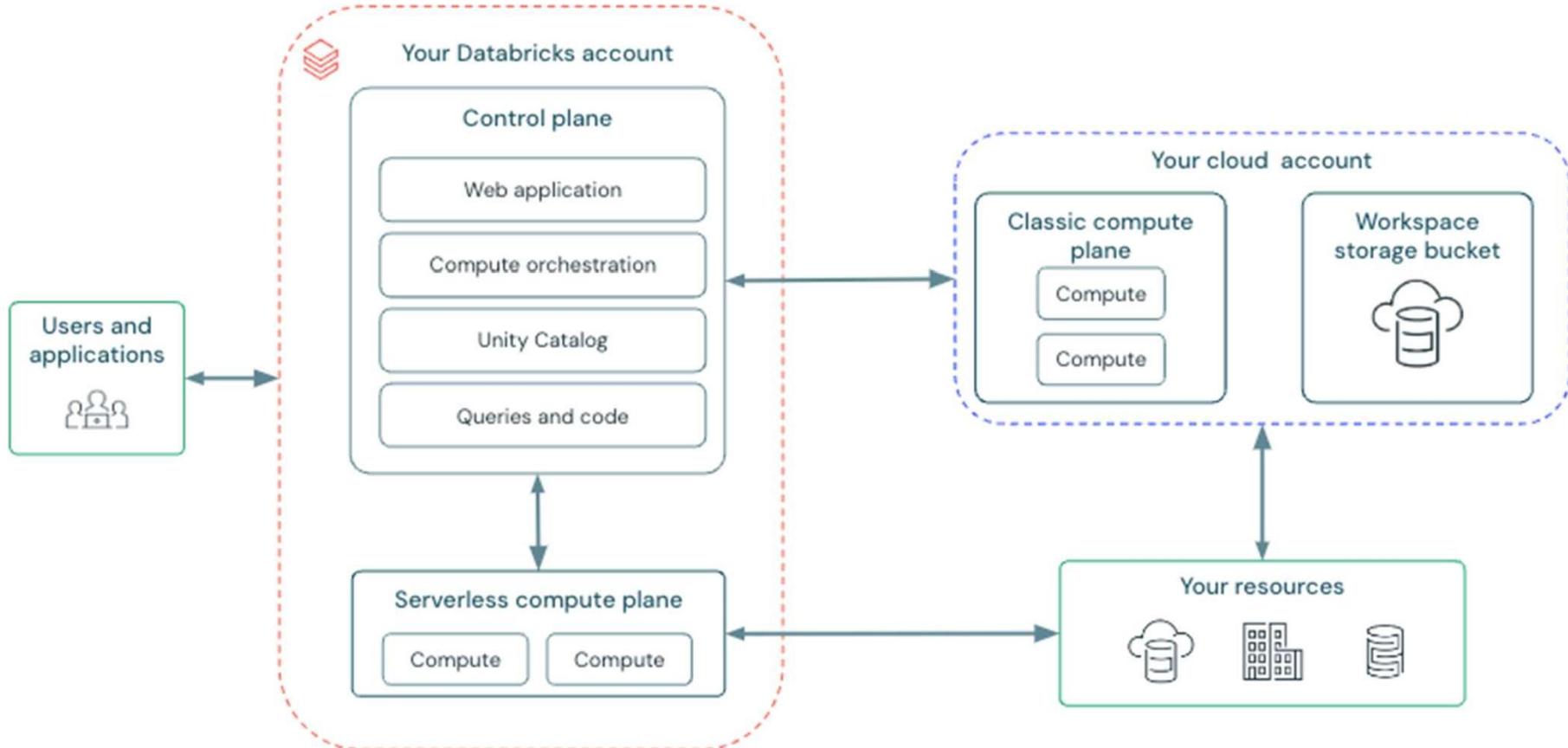
Gobernanza

¿Qué es Databricks & Lakehouse?

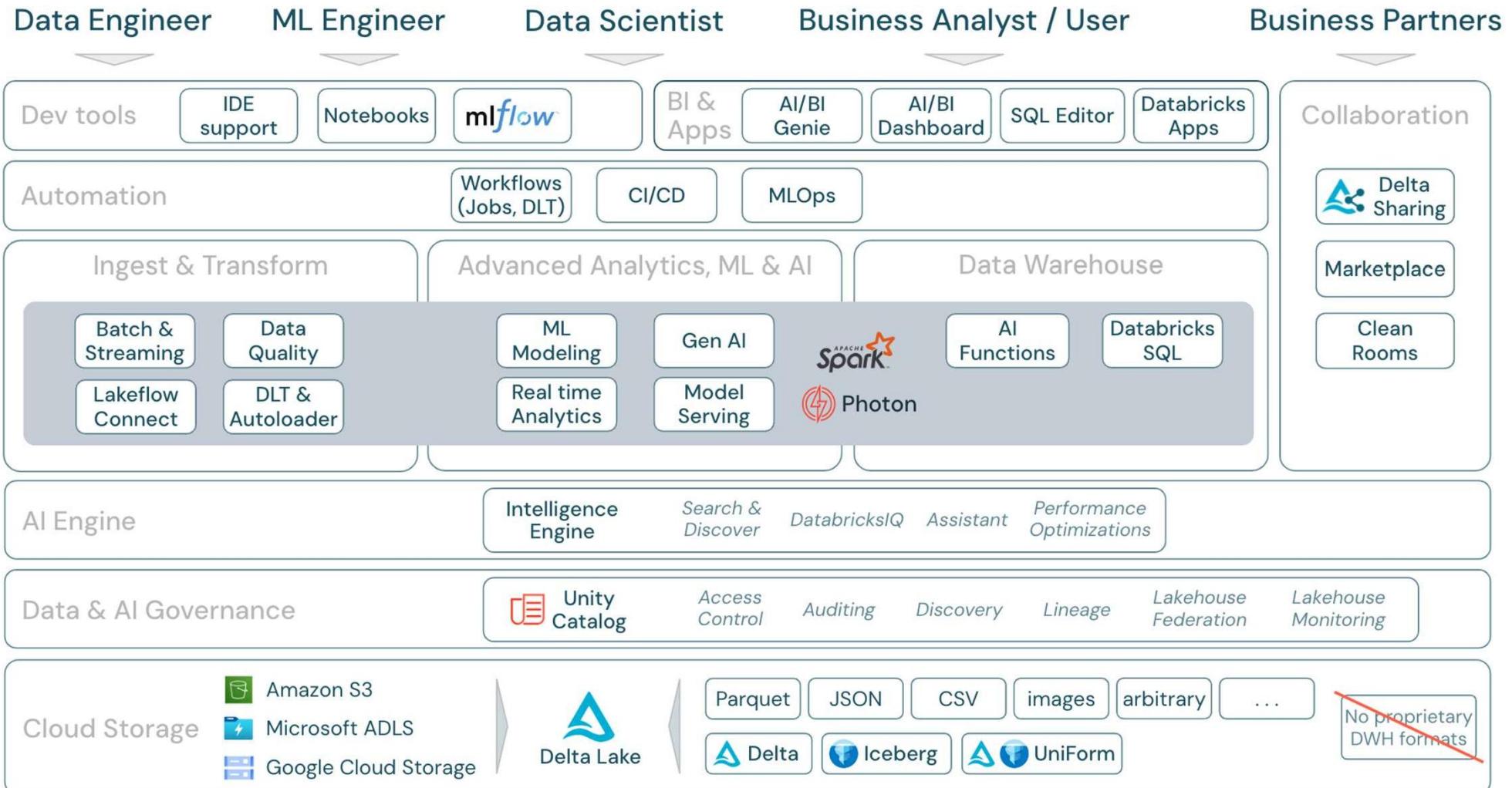


- Unificación de Plataforma de Data
- No vendor locking. Funcionamiento sobre AWS, Azure o Google Cloud.
- Uso de formatos OpenSource. Parquet.
- Uso de Delta Lake (próximas sesiones)

Databricks Architecture (I)



Databricks Architecture (II)



Clusters - Modos de acceso



Single User Access Mode



Shared Access Mode



No Isolation Shared Access Mode

Clusters - Tipos



All-Purpose



Job Clusters



Vector Search Clusters



SQL Warehouses



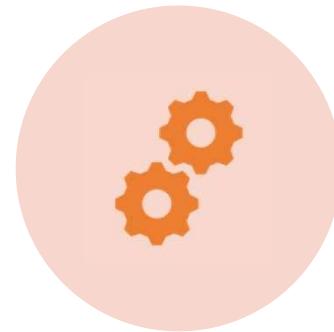
Clusters - Buenas prácticas



DEFINICIÓN PROPÓSITO
DEL CLUSTER



DEFINICIÓN ROLES Y
PERMISOS DE ACCESO Y
USO.



CONFIGURAR AUTO-
TERMINACIÓN

Clusters pools



RECURSO
ADMINISTRADO.



INSTANCIAS
PRECONFIGURADAS
Y READY-TO-USE.



REDUCEN TIEMPOS
DE ESPERA.



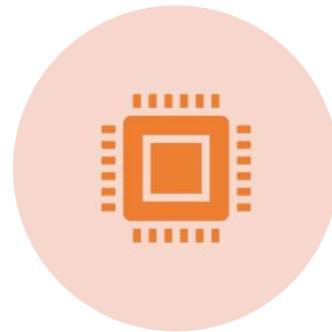
PROPORCIONAN
AGILIDAD Y
ESCALABILIDAD



¿Cuándo utilizar Cluster Pools?



ENTORNO
COLABORATIVO

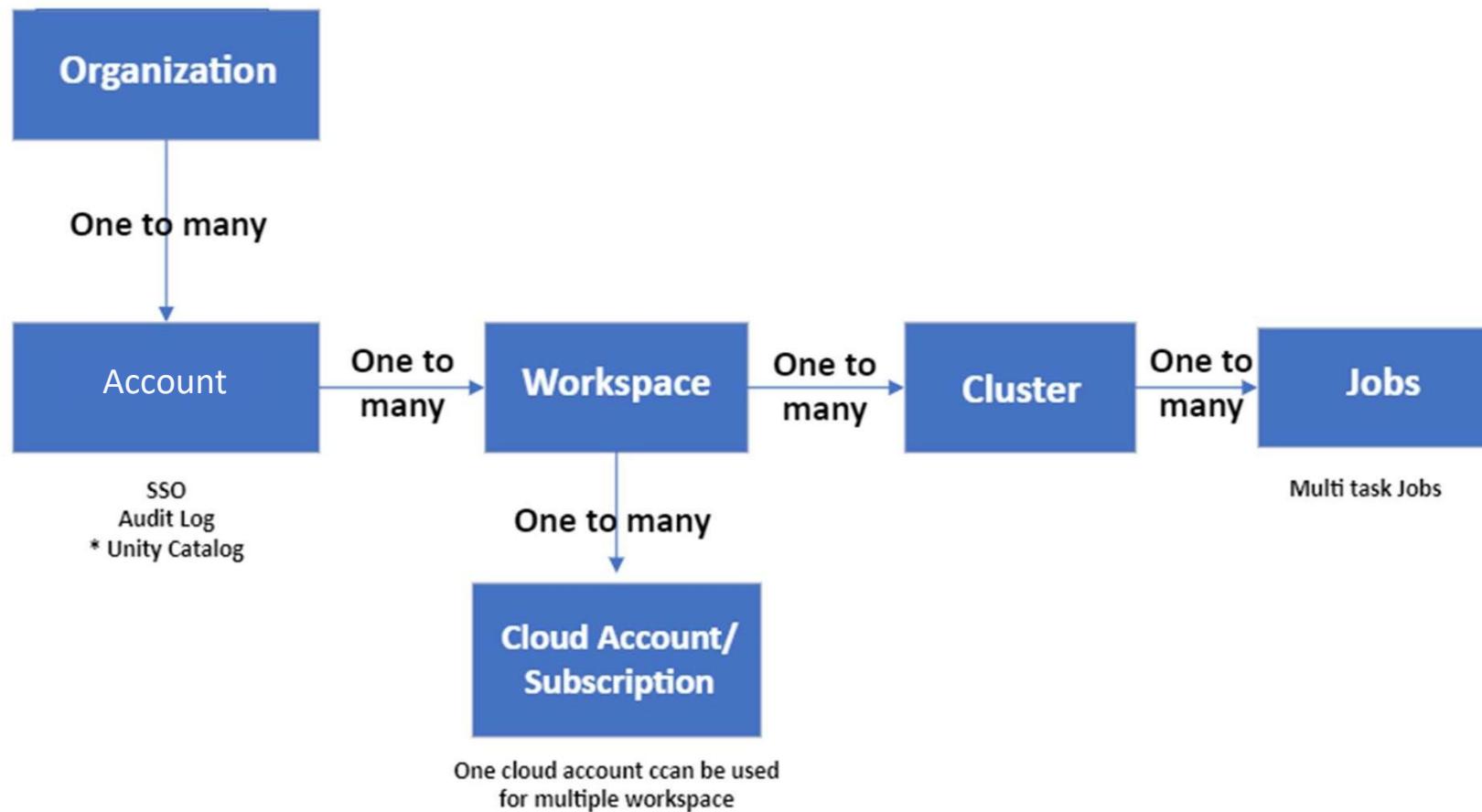


PROCESAMIENTO
ON-DEMAND



OPTIMIZACIÓN
DE COSTES

Workspace



Demo Time

Creación de cuenta en Databricks, configuración de clúster y navegación por el entorno.

[Free Edition | Replacing Databricks Community Edition](#)



Próxima sesión



¿Qué es Spark? Fundamentos.



Spark en Databricks



Ejercicios básicos de Spark, Dataframes y SparkSQL.



Creación de catálogos, esquemas y tablas.

Preguntas / Sugerencias



www.ceste.es