



databricks

La **explosiva** pareja del análisis Big Data

Eugenio Marinetto

27 abril 2023

PyData Granada



Bienvenida



Nombre



Perfil Tech



Experiencia
Python



Expectativas
del taller

Empiezo yo



Nombre



Perfil Tech



Experiencia
Python



Expectativas del taller



A Eugenio Marinetto PhD
Torre del Mar
Málaga

Teleco por la UGR
PhD en Bioingeniería por la UC3M

Actualmente trabajando como:

Científico de Datos/MLOps

Espero que aprendáis algo!

Yours sincerely,
Minerva McGonagall
Deputy Headmistress



Historia de la Magia



- △ Librería para procesamiento **Big Data**
- △ Proyecto **open-source** más grande en procesamiento de datos
- △ De los fundadores de **Databricks**

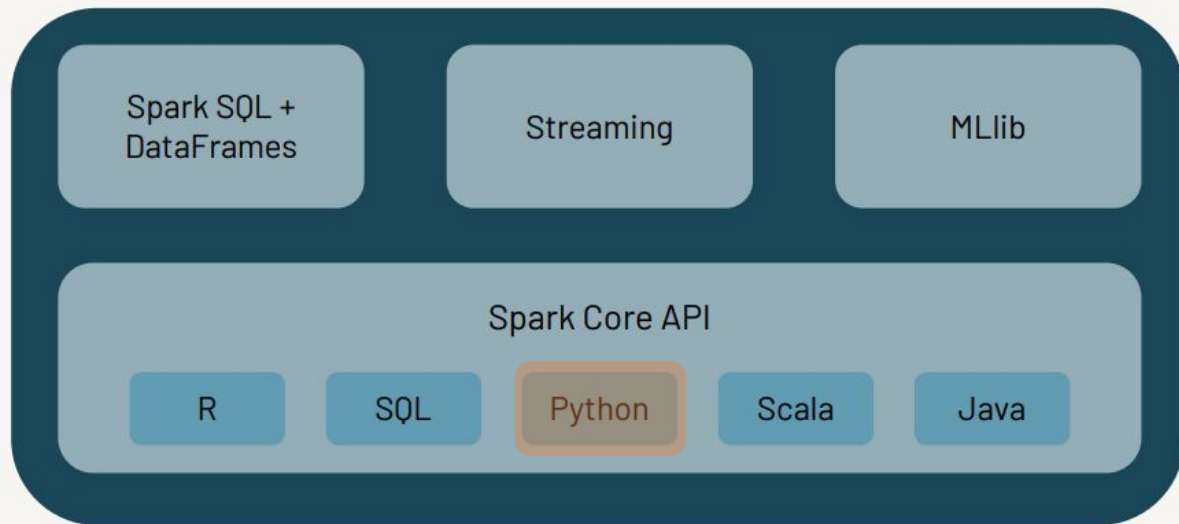




Historia de la Magia

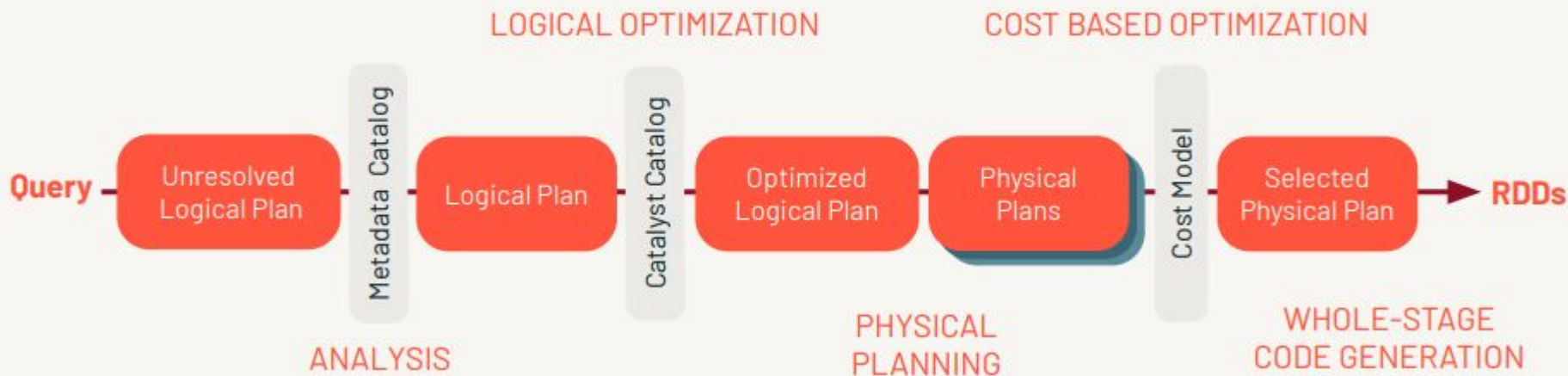


Spark API





Optimización de *Queries*





Historia de la Magia



databricks



Plataforma Cloud para el procesamiento
Big Data



Multicloud



Orquestación de Pipelines de
procesamiento

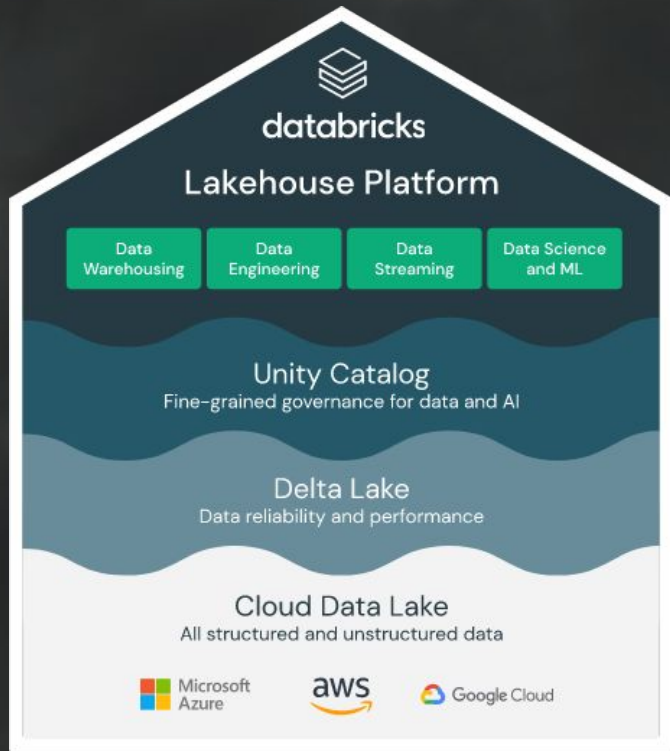




Historia de la Magia



databricks





Astronomía



Astronomía

1. Entrar a Databricks y primer contacto
2. Levantar el primer cluster
3. Crear el primer notebook
4. Probar y ejecutar





Encantamientos



Encantamientos y transformaciones

1. Subir datos y crear tablas (Metastore)
2. La función display
3. Procesamiento básico con PySpark
4. Ventanas !
5. Guardar y persistir





Vuelo





Vuelo

1. Primer pipeline
2. Parámetros de un notebook
3. Programar un workflow





Defensa contra las artes oscuras



Defensa contra las artes oscuras

1. UDFs
2. Particionado
3. Formato Delta



Gracias Muggles

