Introducción a Databricks y el ecosistema Spark

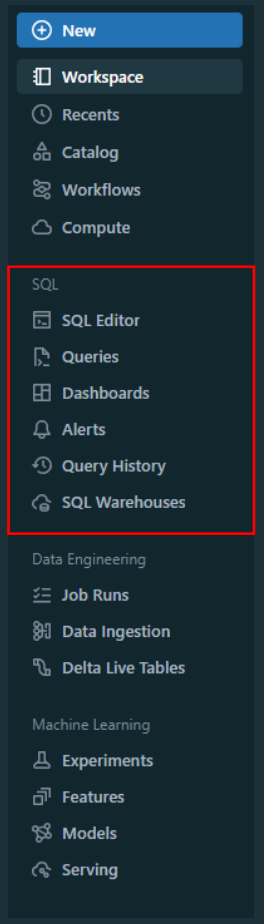
Databricks

Data Analytics

databricks

IMMUNE
TECHNOLOGY INSTITUTE

Databricks – Data Analytics



Databricks ofrece un ecosistema unificado para el análisis de datos, permitiendo a los usuarios consultar datos ingestados a través de SQL y una interfaz gráfica intuitiva.

La plataforma combina la facilidad de SQL con las herramientas avanzadas de análisis para un procesamiento de datos eficaz, habilitando desde la exploración de datos hasta la creación de Dashboards y Alertas.

databricks

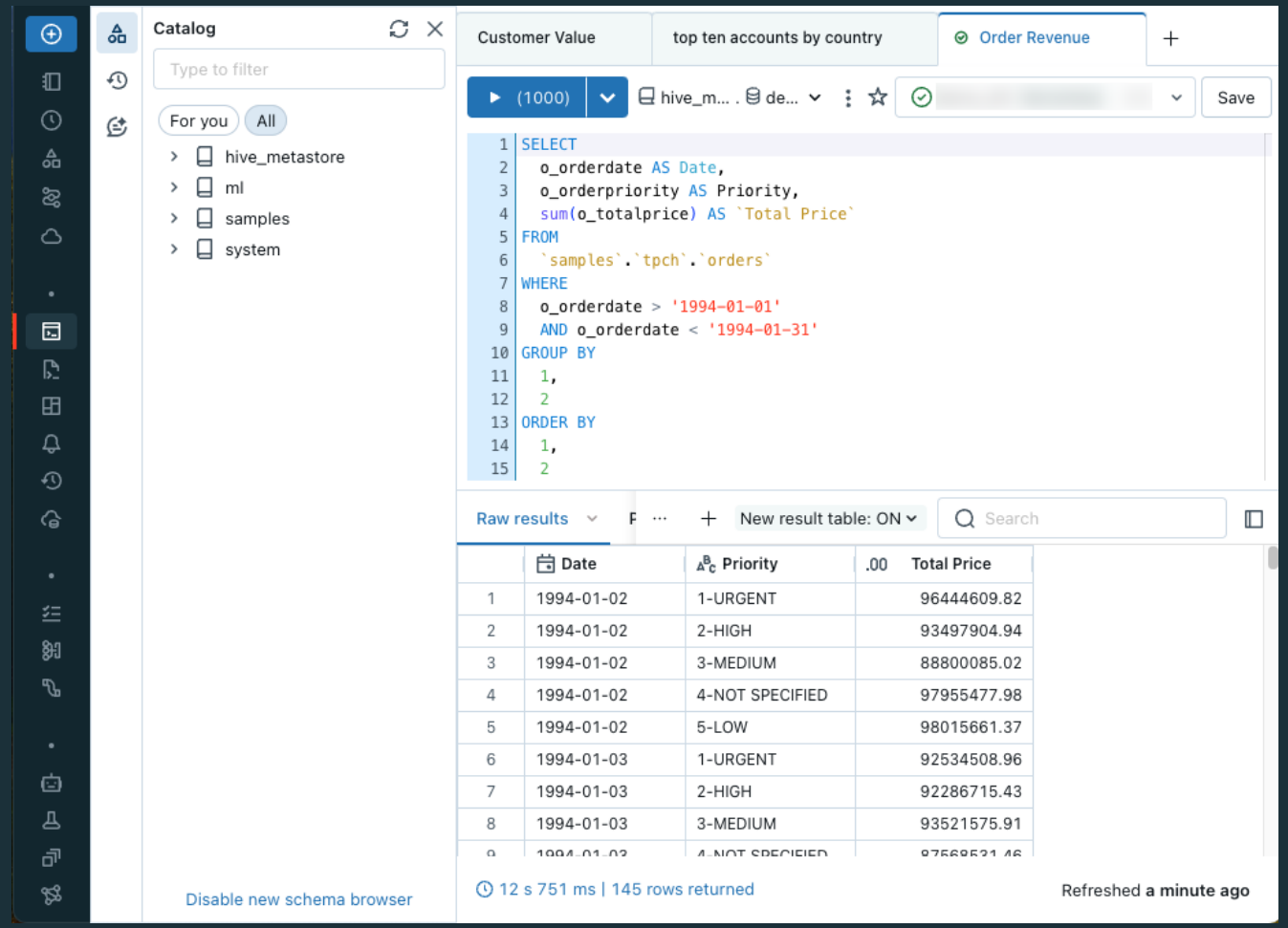
IMMUNE
TECHNOLOGY INSTITUTE

Consulta de Datos con SQL

Utilizando SQL en Databricks, los usuarios pueden ejecutar consultas sobre los datos ingesados de manera eficiente.

La sección 'SQL Editor' dentro del menú de opciones permite escribir y ejecutar consultas SQL directamente, facilitando la manipulación y análisis de grandes volúmenes de datos.

Para consultas que se repitan con frecuencia podemos guardarlas para recuperlas posteriormente a través de la opción “Queries”



databricks

IMMUNE TECHNOLOGY INSTITUTE

SQL Warehouses

Los SQL Warehouses representan una evolución en la gestión y análisis de datos en la nube.

Diseñados específicamente para consultas SQL de alto rendimiento, estos clusters están optimizados para cargar, transformar y analizar grandes volúmenes de datos con una eficiencia excepcional

Se diferencian de los clusters de procesamiento de datos en las siguientes características:

- **Optimización para Consultas SQL**: Mediante una cola de ejecución en paralelo y Photon por defecto habilitado para la ejecución de consultas
- **Escalabilidad Automática**: Escalado automático para ajustarse a la carga de trabajo
- **Gestión de Costes**: Más predecible basado en el rendimiento de las consultas.
- **Experiencia de Usuario y Herramientas de BI**: Están diseñados para ser integrados con herramientas de BI. Con los clusters normales, la experiencia no es tan directa ni optimizada como con SQL Warehouses

SQL Warehouses en Azure

Databricks ofrece una variedad de tamaños de clusters, desde pequeños para tareas ligeras de desarrollo hasta grandes para el procesamiento de datos a escala de petabytes.

Esto permite a los usuarios escalar sus recursos según las necesidades específicas de cada proyecto.

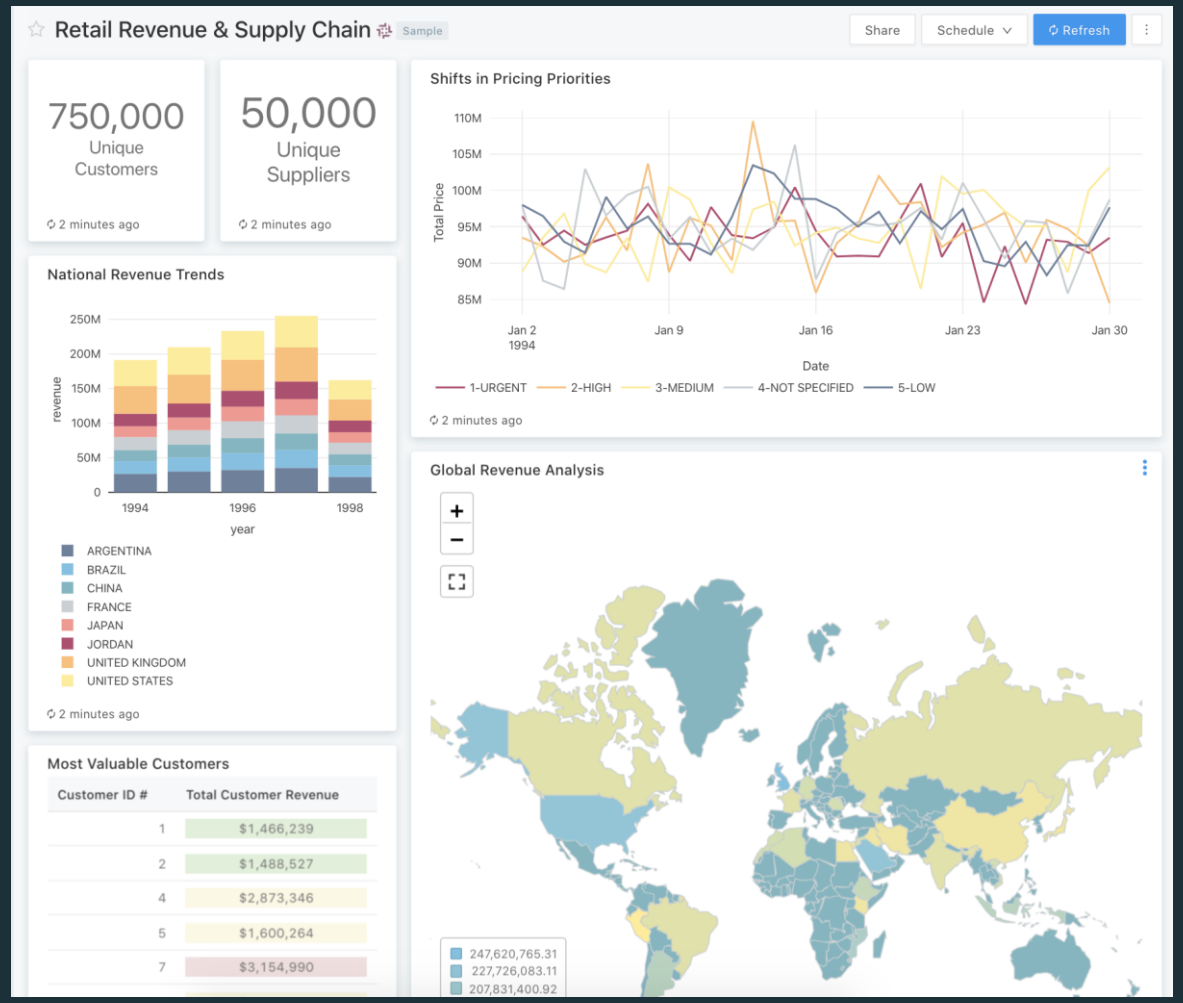
Cluster size	Instance type for driver (applies only to pro and classic SQL warehouses)	Worker count
2X-Small	Standard_E8ds_v4	1 x Standard_E8ds_v4
X-Small	Standard_E8ds_v4	2 x Standard_E8ds_v4
Small	Standard_E16ds_v4	4 x Standard_E8ds_v4
Medium	Standard_E32ds_v4	8 x Standard_E8ds_v4
Large	Standard_E32ds_v4	16 x Standard_E8ds_v4
X-Large	Standard_E64ds_v4	32 x Standard_E8ds_v4
2X-Large	Standard_E64ds_v4	64 x Standard_E8ds_v4
3X-Large	Standard_E64ds_v4	128 x Standard_E8ds_v4
4X-Large	Standard_E64ds_v4	256 x Standard_E8ds_v4

Creación de Dashboards y Visualizaciones Interactivas

Databricks facilita la creación de dashboards y visualizaciones interactivas, permitiendo a los usuarios transformar sus datos en insights accionables.

La plataforma soporta una amplia gama de gráficos y visualizaciones personalizables.

Permite compartir los resultados de manera efectiva con stakeholders, potenciando la toma de decisiones basada en datos



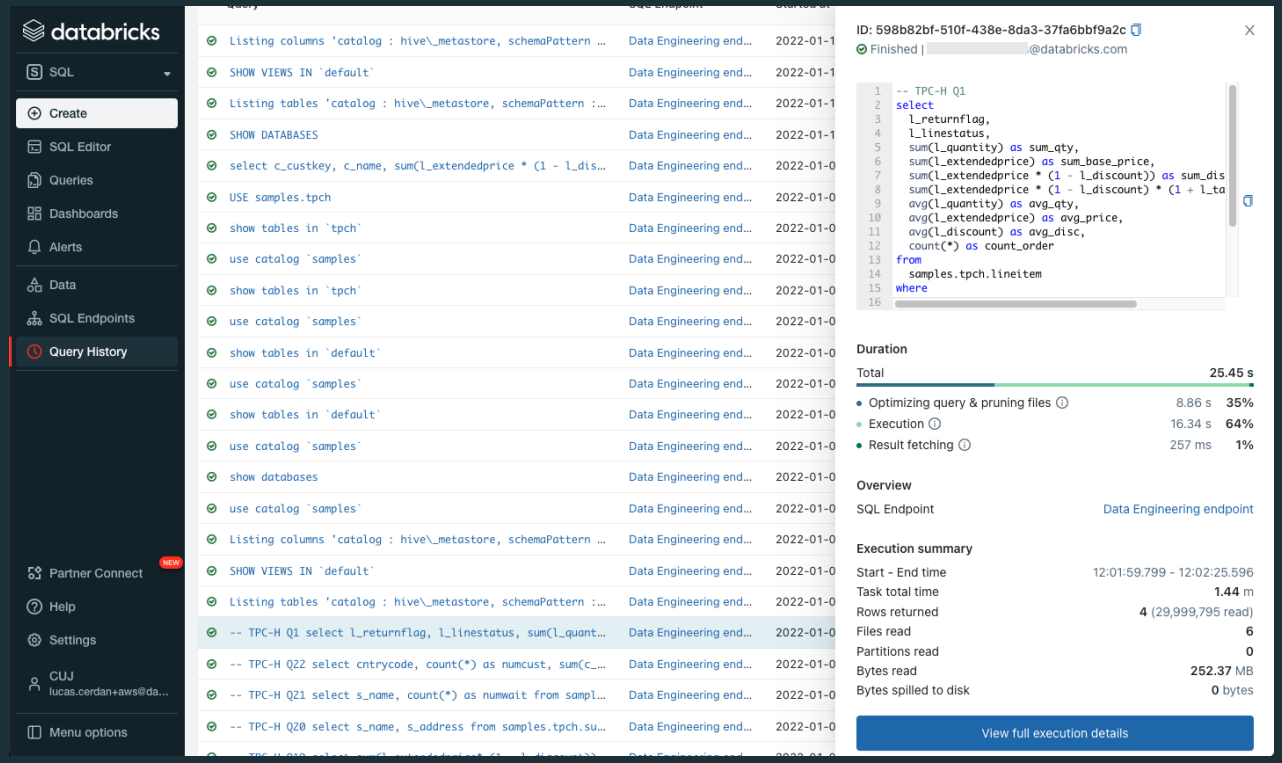
databricks

IMMUNE TECHNOLOGY INSTITUTE

Query History

El Query History es una herramienta vital en Databricks, ya que permite a los usuarios rastrear y analizar el rendimiento de sus consultas SQL.

Esto no solo ayuda en la optimización de consultas sino también en la auditoría y el cumplimiento de políticas de gobernanza de datos.



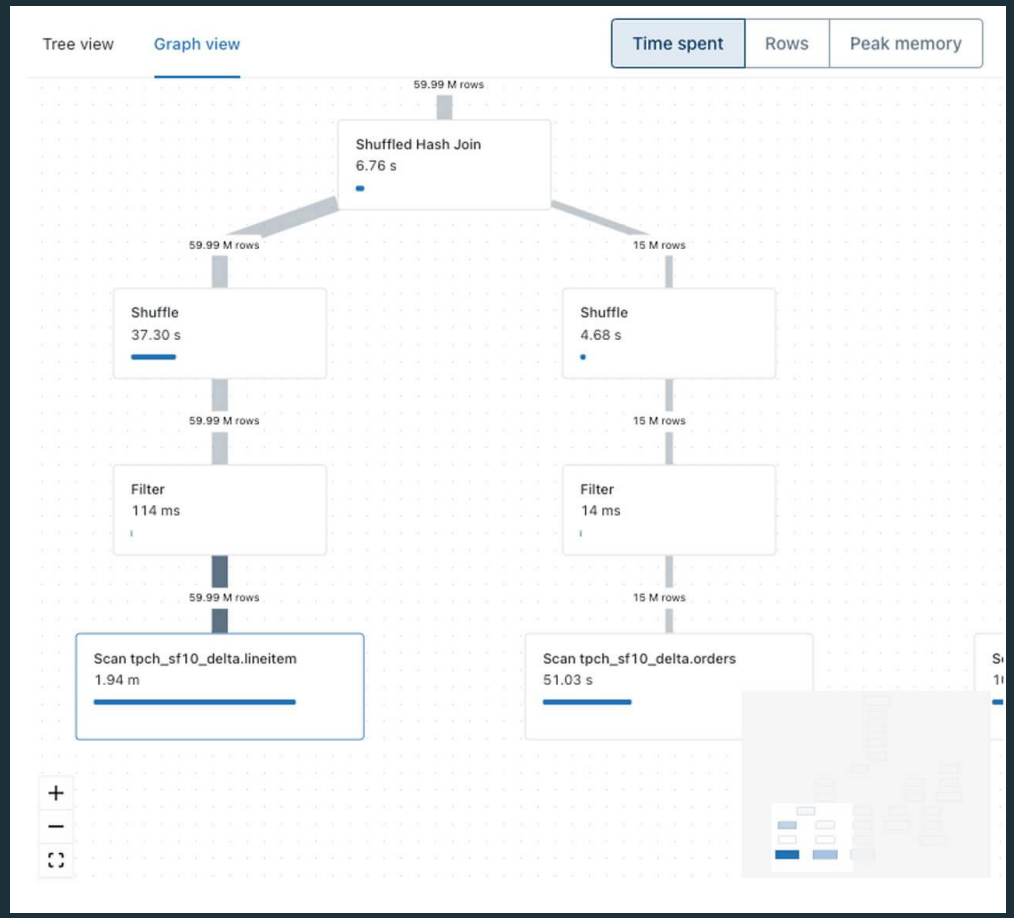
databricks

IMMUNE TECHNOLOGY INSTITUTE

Query History

El Query History es una herramienta vital en Databricks, ya que permite a los usuarios rastrear y analizar el rendimiento de sus consultas SQL.

Esto no solo ayuda en la optimización de consultas sino también en la auditoría y el cumplimiento de políticas de gobernanza de datos.



databricks

IMMUNE TECHNOLOGY INSTITUTE

Integración con Herramientas Externas de BI y Visualización

Databricks se integra a la perfección con herramientas externas de BI y visualización como Tableau, Power BI y Looker, ampliando las capacidades de análisis y reporte de los usuarios.

Esta integración permite a las organizaciones aprovechar sus inversiones existentes en herramientas de BI, facilitando la visualización avanzada de datos y la creación de dashboards interactivos

databricks

Lab 3: Databricks Data Analytics

IMMUNE
TECHNOLOGY INSTITUTE