Taller 2

Daniel Guillermo Crespo Duarte – Código: 40573

UNIVERSIDAD ECCI

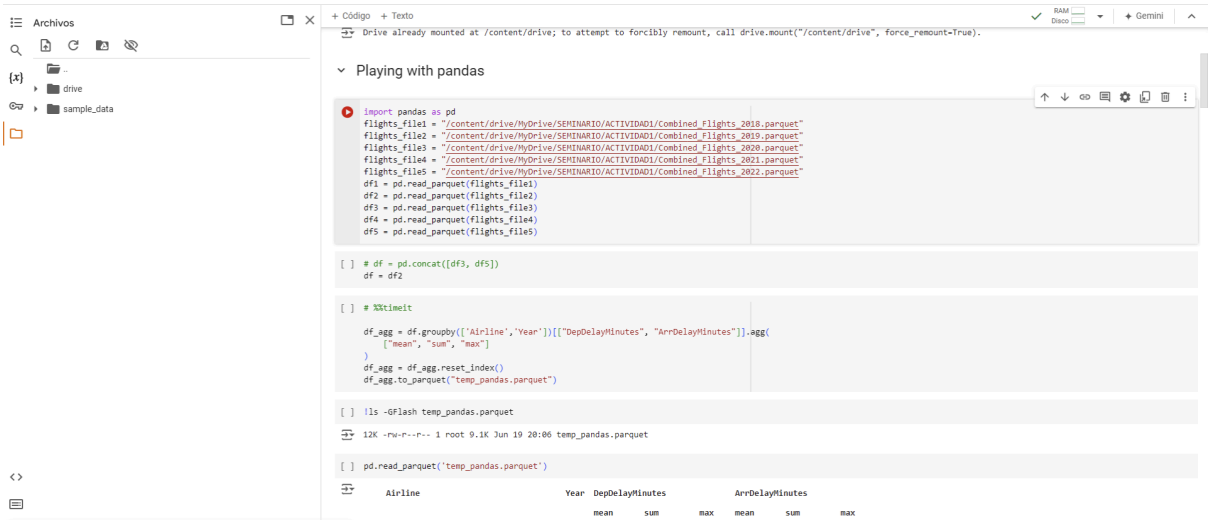
SEMINARIO BIG DATA Y GERENCIA DE DATOS
2024-1

PROFESOR : ELIAS BUITRAGO BOLIVAR
BOGOTÁ, D.C.

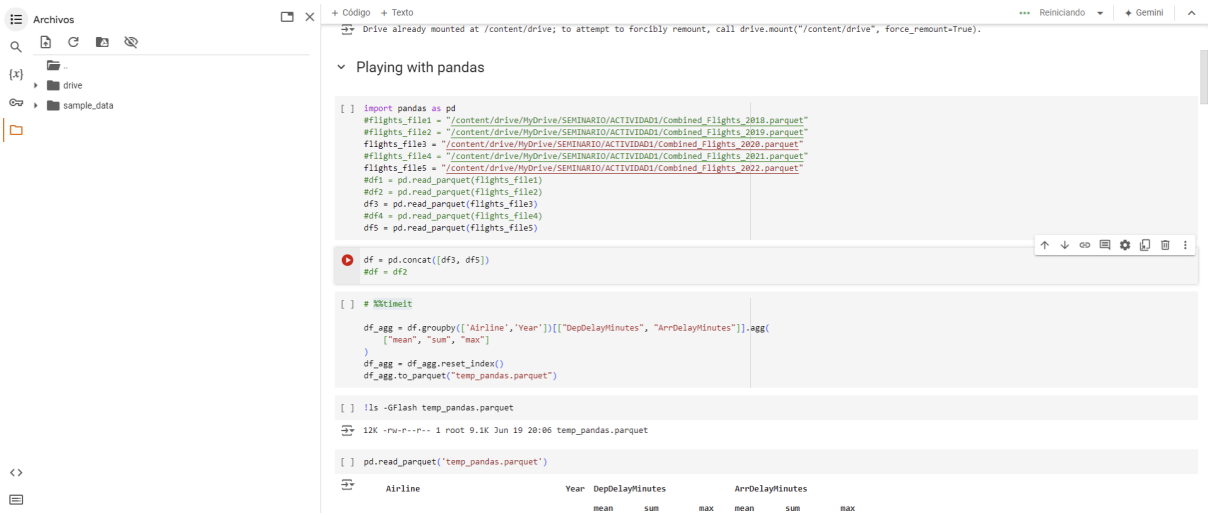
2024

Resultados de Pandas

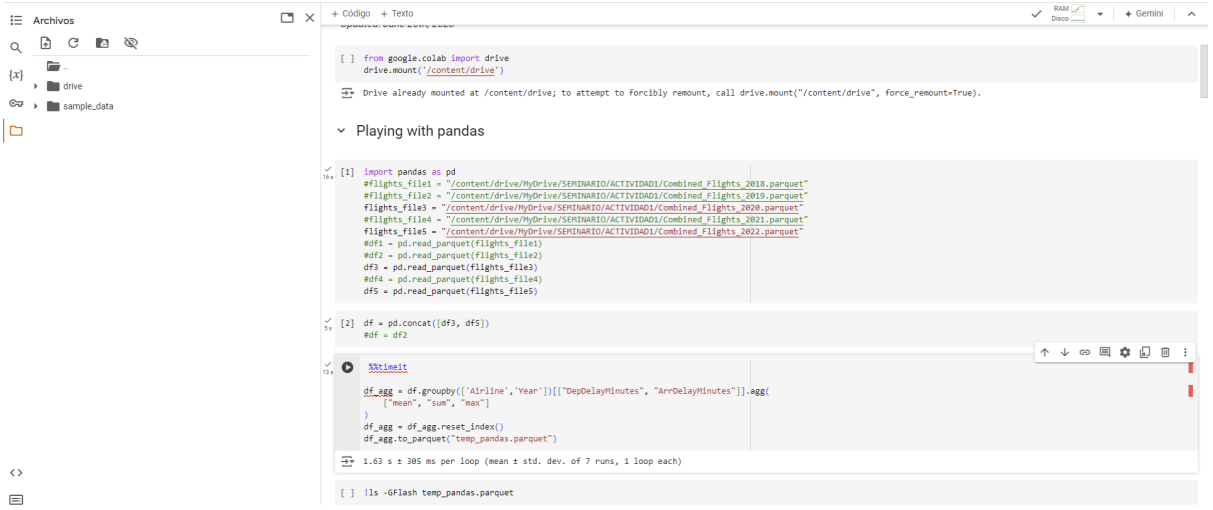
Importar archivos:



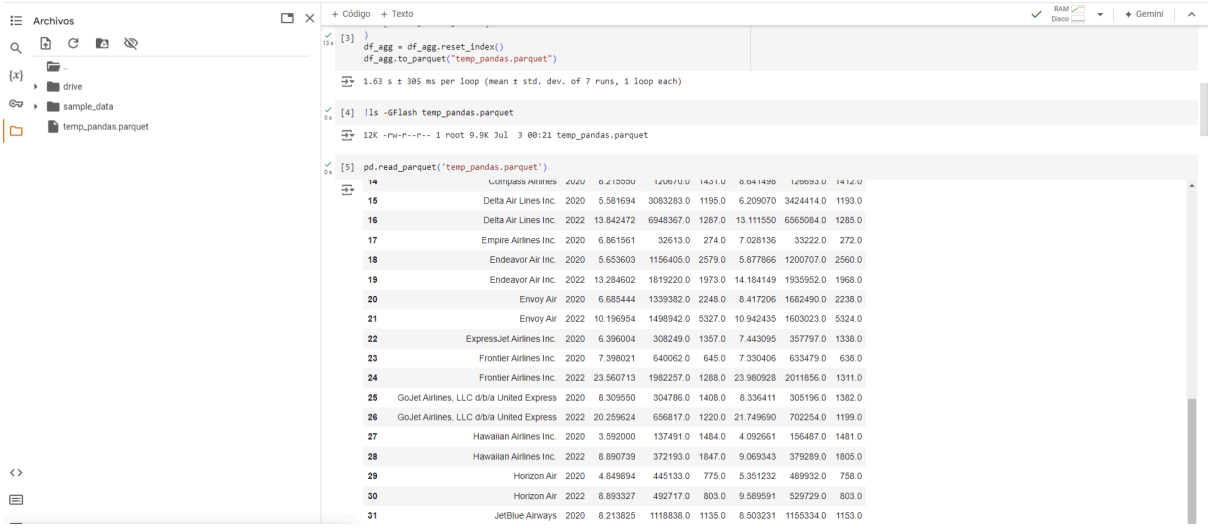
Realizar pruebas con los archivos 3 y 5:



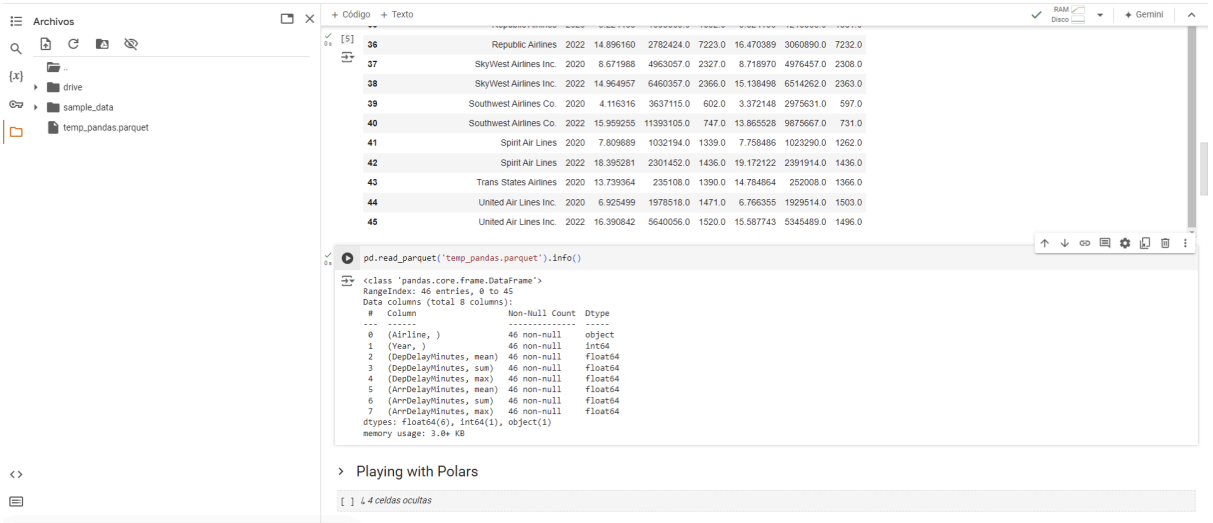
Obtener información de los archivos:



Resultados del modelo:

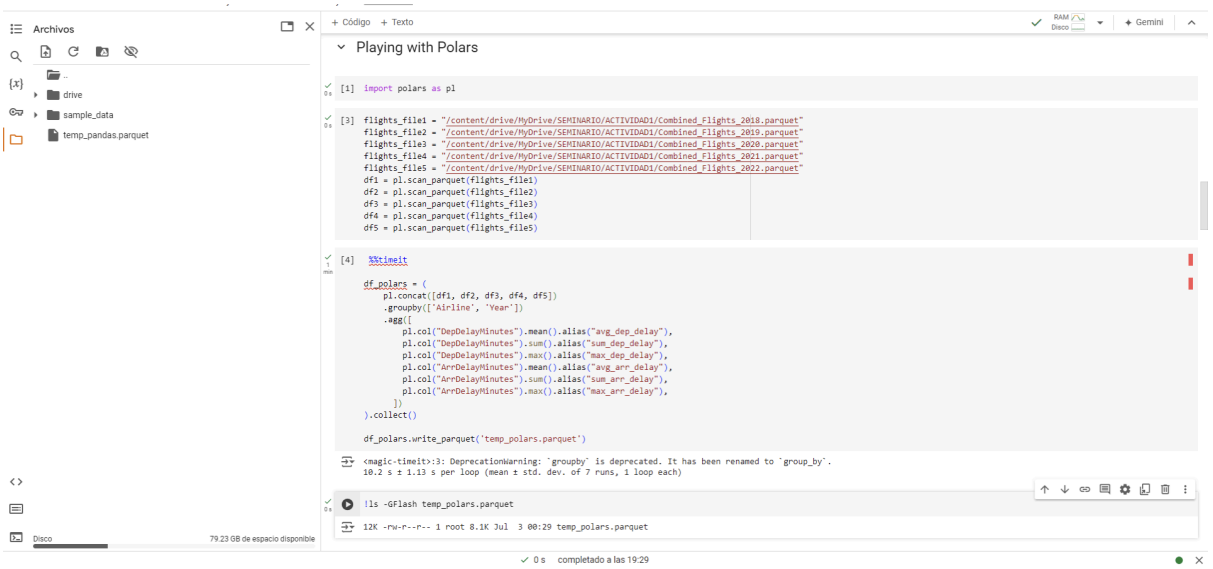


15	Delta Air Lines Inc.	2020	5.581694	3083283.0	1195.0	6.209070	3424414.0	1193.0
16	Delta Air Lines Inc.	2022	13.842472	6948367.0	1287.0	13.111550	6565084.0	1285.0
17	Empire Airlines Inc.	2020	6.861561	32613.0	274.0	7.028136	33222.0	272.0
18	Endeavor Air Inc.	2020	5.653603	1156405.0	2579.0	5.877866	1200707.0	2560.0
19	Endeavor Air Inc.	2022	13.284602	1819220.0	1973.0	14.184149	1935952.0	1968.0
20	Envoy Air	2020	6.685444	1339382.0	2248.0	8.417206	1682490.0	2238.0
21	Envoy Air	2022	10.196954	1498942.0	5327.0	10.942435	1603023.0	5324.0
22	ExpressJet Airlines Inc.	2020	6.396004	308249.0	1357.0	7.443095	357797.0	1338.0
23	Frontier Airlines Inc.	2020	7.398021	640062.0	645.0	7.330406	633479.0	638.0
24	Frontier Airlines Inc.	2022	23.560713	1982257.0	1288.0	23.980928	2011856.0	1311.0
25	GoJet Airlines, LLC d/b/a United Express	2020	8.309550	304786.0	1408.0	8.336411	305196.0	1382.0
26	GoJet Airlines, LLC d/b/a United Express	2022	20.259624	656817.0	1220.0	21.749690	702254.0	1199.0
27	Hawaiian Airlines Inc.	2020	3.592000	137491.0	1484.0	4.092661	156487.0	1481.0
28	Hawaiian Airlines Inc.	2022	8.890739	372193.0	1847.0	9.069343	379289.0	1805.0
29	Horizon Air	2020	4.849894	445133.0	775.0	5.351232	489932.0	758.0
30	Horizon Air	2022	8.893327	492717.0	803.0	9.589591	529729.0	803.0
31	JetBlue Airways	2020	8.213825	1118838.0	1135.0	8.503231	1155334.0	1153.0

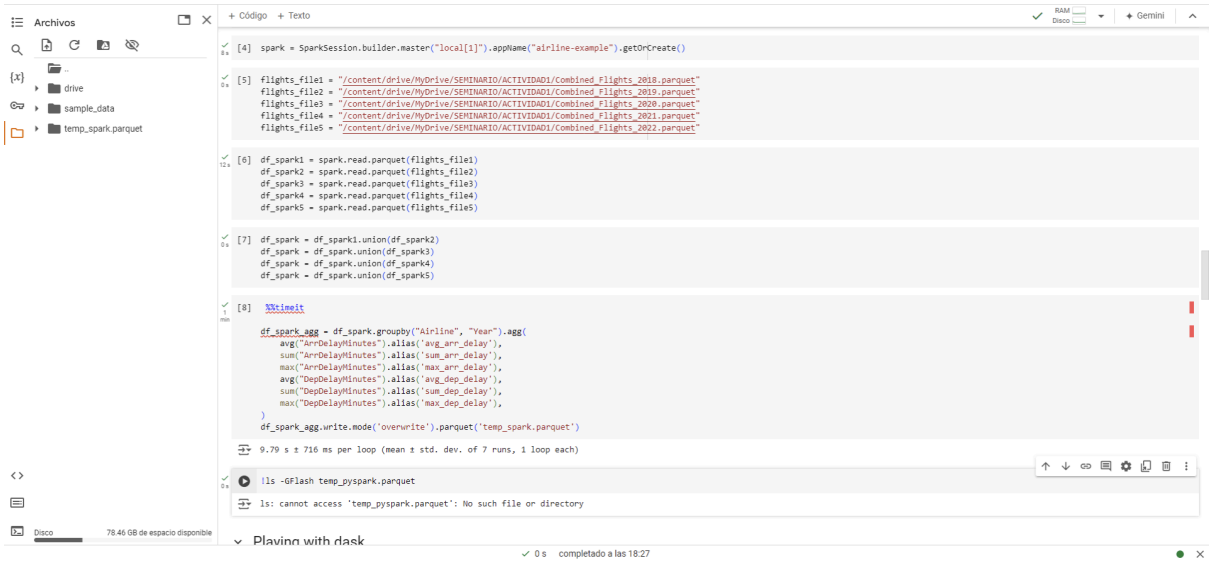


36	Republic Airlines	2022	14.896160	2782424.0	7223.0	16.470389	3060890.0	7232.0
37	SkyWest Airlines Inc.	2020	8.671988	4963057.0	2327.0	8.718970	4976457.0	2308.0
38	SkyWest Airlines Inc.	2022	14.964957	6460357.0	2366.0	15.138498	6514262.0	2363.0
39	Southwest Airlines Co.	2020	4.116316	3637115.0	602.0	3.372148	2975631.0	597.0
40	Southwest Airlines Co.	2022	15.959255	11393105.0	747.0	13.865528	9875667.0	731.0
41	Spirit Air Lines	2020	7.809689	1032194.0	1339.0	7.758486	1023290.0	1262.0
42	Spirit Air Lines	2022	18.395281	2301452.0	1436.0	19.172122	2391914.0	1436.0
43	Trans States Airlines	2020	13.739364	235108.0	1390.0	14.784864	252008.0	1366.0
44	United Air Lines Inc.	2020	6.925499	1978518.0	1471.0	6.766355	1929514.0	1503.0
45	United Air Lines Inc.	2022	16.390842	5640056.0	1520.0	15.587743	5345489.0	1496.0

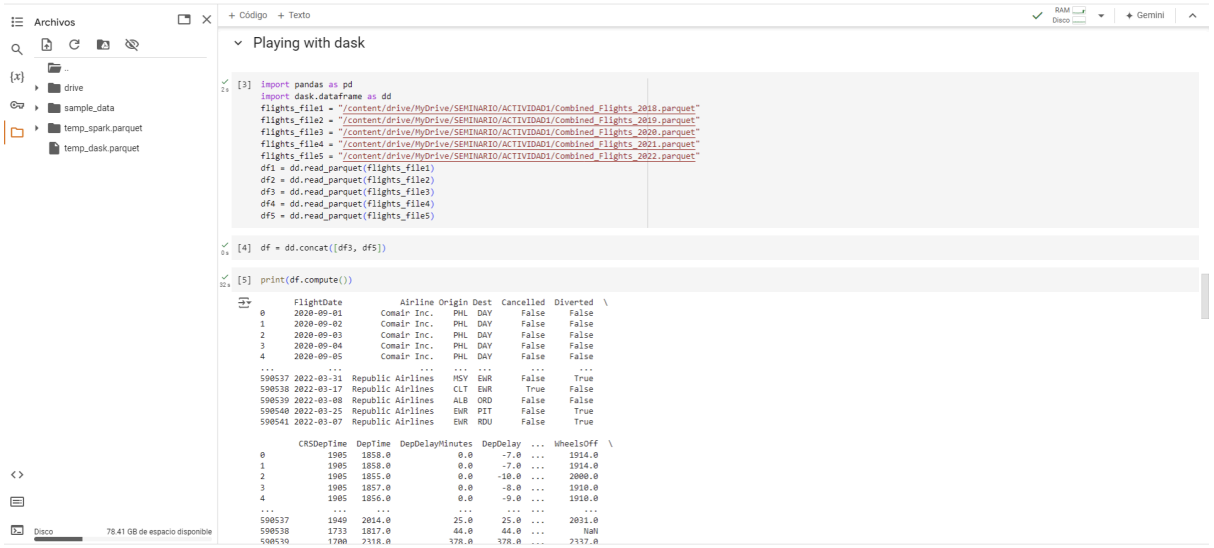
Resultados de Polars



Resultados de Pyspark



Resultados de Dask



Comparación

- Pandas no puede analizar 5 datos a la vez, solo puede procesar 2 y solo permite agrupar de a 2 por lo cual no toma tanto tiempo su ejecución.
- Polaris sí puede analizar 5 datos a la vez, y puede agrupar 5 datos pero su tiempo de ejecución es más extenso.
- PySpark si puede analizar 5 datos a la vez, pero no puede agrupar datos, sólo puede analizarlos en paralelo y tiene un rendimiento similar a Polaris.
- Dask si puede analizar 5 datos a la vez, pero no puede agrupar datos, sólo puede analizarlos en paralelo.