

Analisis de conversion de archivos con y sin compresión

Rodrigo Ignacio Ramírez Díaz

April 19, 2025

1 Analisis temporal de conversion y compresion de arcvhivos

Para el análisis temporal se cuenta con la siguiente 1, la cual contiene el tiempo que se demora en transformar los archivos CSV a formato AVRO o Parquet, teniendo com variacion el porcentaje de datos a transformar del dataset y original y si la conversion es o no con algun formato de compresión.

% datos	Avro			Parquet			
	Sin comprimir	Deflate	Snappy	Sin comprimir	Snappy	Gzip	LZ4
1%	0,08	0,04	0,05	0,09	0,06	0,04	0,06
10%	0,79	0,41	0,52	0,67	0,46	0,37	0,48
20%	1,97	1,03	1,29	1,40	1,02	0,83	1,06
50%	3,94	2,05	2,58	2,45	1,89	1,57	1,96
100%	7,88	3,48	4,63	4,06	3,14	2,49	3,15

Table 1: Comparación de tamaños de archivo (MB) en Avro y Parquet bajo distintos porcentajes de datos y métodos de compresión.

1.1 Formato de archivo CSV a AVRO

1.2 Formato de archivo CSV a Parquet

2 Preguntas

1. ¿Qué conclusiones puede obtener de los resultados anteriores?
ondocsndc
2. Basado en los resultados: ¿Qué combinación (formato/compresión) elegiría para almacenar el dataset en un data lake en la nube? Justifique su respuesta.

3. ¿Cual fue el principal desafio para desarrollar la presente tarea?