

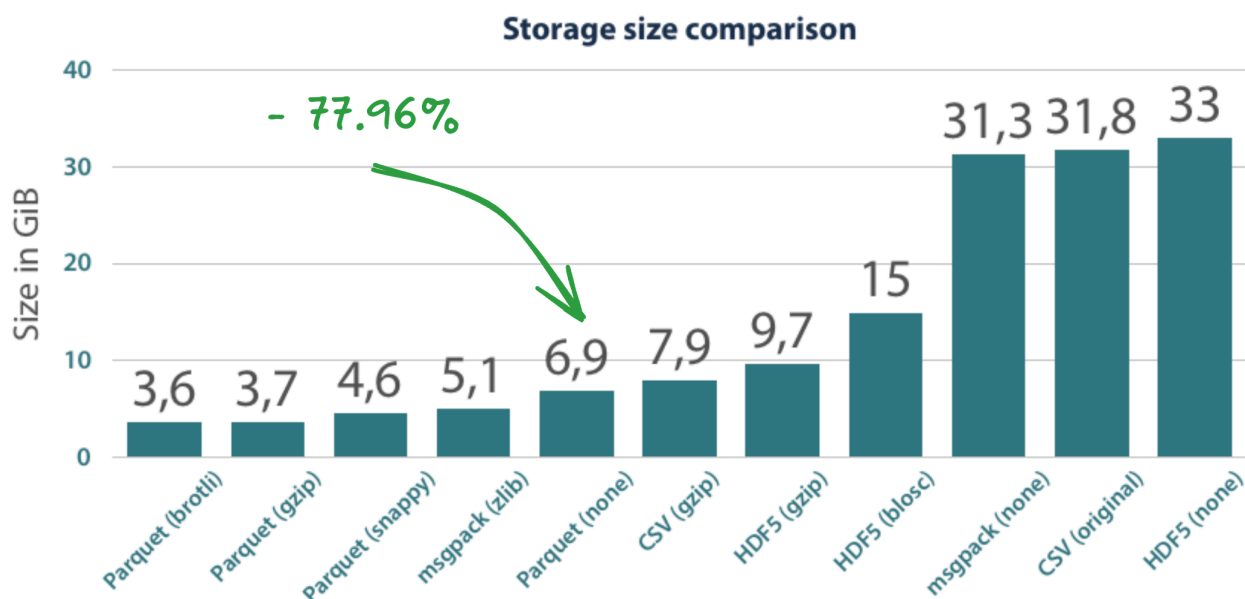
007 USO DE PARQUET



En el mundo de los datos, los formatos son esenciales para almacenar y compartir información. Los más comunes incluyen archivos como CSV, TXT, JSON, XML y Excel. Sin embargo, en la analítica y la ingeniería de datos, estos formatos suelen ser ineficientes en términos de almacenamiento y rendimiento.

Parquet, en cambio, es un formato columnar optimizado para grandes volúmenes de datos. Es altamente eficiente en espacio y velocidad, permitiendo consultas rápidas y almacenamiento compacto. Su capacidad para manejar datos a gran escala e integrarse con tecnologías de Big Data como Hadoop y Spark lo ha convertido en el formato dominante en la actualidad.

Miremos una comparativa:



DuckDB tiene una característica única que no poseen otras bases de datos: permite leer directamente los archivos y utilizar todas las funciones de la base de datos. Además, soporta los formatos mencionados anteriormente, excepto XML.

LEER ARCHIVOS

Para leer datos de un archivo Parquet, utilice la función `read_parquet` en la cláusula `FROM` de una consulta:

```
-- forma #1  
SELECT * FROM read_parquet("ruta_local_archivo")
```

```
-- forma #2  
SELECT * FROM "ruta_local_archivo"
```

EXPORTAR ARCHIVOS

Para exportar los datos de una tabla a un archivo Parquet, utilice la declaración `COPY`:

```
COPY tabla TO 'ruta/nombre_archivo.parquet' (FORMAT PARQUET);
```

El resultado de las consultas también se puede exportar directamente a un archivo Parquet:

```
COPY (SELECT col_1, col_n FROM tabla) TO 'ruta/nombre_archivo.parquet'  
(FORMAT PARQUET);
```

Para más información y detalles sobre su uso, pueden consultar el siguiente enlace:

[duckdb-formatos](#)