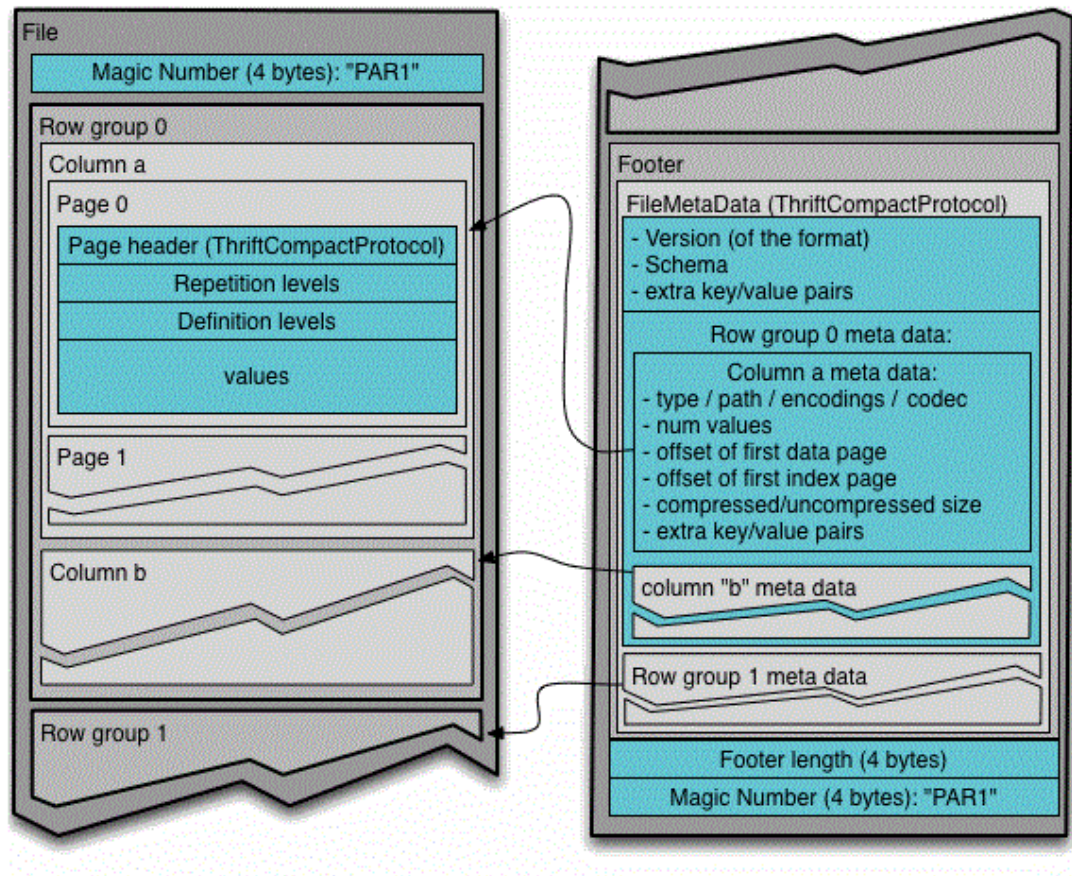


# Formato: Parquet

El formato Parquet es un formato open-source de almacenamiento en columnas enfocado en la mejora del procesamiento de datos, modelado de datos y programación



El formato de Parquet está compuesto por tres piezas:

- **Row group:** es un conjunto de filas en formato columnar, con un tamaño entre 50Mb a 1Gb.
- **Column chunk:** son los datos de una columna en un grupo. Se puede leer de manera independiente para mejorar las lecturas.
- **Page:** Es donde finalmente se almacenan los datos debe ser lo suficiente grande para que la compresión sea eficiente.

Rendimiento de la consulta

Dependiendo del caso de uso de su negocio, Apache Parquet es una buena opción **si tiene que proporcionar funciones de búsqueda parcial, es decir, no consultar todas las columnas**, y no le preocupa el tiempo de escritura del archivo.