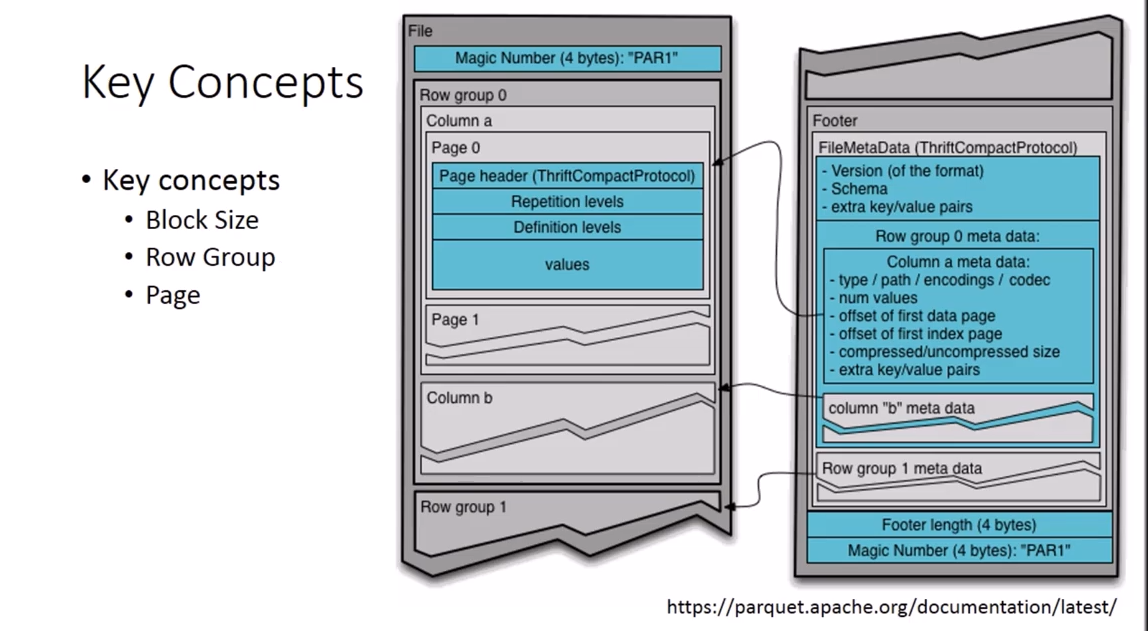
**Parquet:**

* Block (hdfs block): This means a block in hdfs and the meaning is unchanged for describing this file format. The file format is designed to work well on top of hdfs.
* File: A hdfs file that must include the metadata for the file. It does not need to actually contain the data.
* Row group: A logical horizontal partitioning of the data into rows. There is no physical structure that is guaranteed for a row group. A row group consists of a column chunk for each column in the dataset.
* Column chunk: A chunk of the data for a particular column. These live in a particular row group and is guaranteed to be contiguous in the file.
* Page: Column chunks are divided up into pages. A page is conceptually an indivisible unit (in terms of compression and encoding). There can be multiple page types which is interleaved in a column chunk.

Hierarchically, a file consists of one or more row groups. A row group contains exactly one column chunk per column. Column chunks contain one or more pages.



Reference1: <https://www.youtube.com/watch?v=rVC9F1y38oU&t=88s>

Reference2: <http://parquet.apache.org/documentation/latest/>