Compared with RCFile format, for example, ORC file format has many advantages such as:

* a single file as the output of each task, which reduces the NameNode’s load
* Hive type support including datetime, decimal, and the complex types
* light-weight indexes stored within the file
* skip row groups that don’t pass predicate filtering
* seek to a given row
* block-mode compression based on data type
* run-length encoding for integer columns
* dictionary encoding for string columns
* concurrent reads of the same file using separate RecordReaders
* ability to split files without scanning for markers
* bound the amount of memory needed for reading or writing
* metadata stored using Protocol Buffers, which allows addition and removal of fields