Explain File Formats in Hive in Brief with an example.

Different file formats that are used in hive are

1. TEXTFILE FORMAT

This is format is basic and famous file format and it can load data of from CSV (Comma Separated Values), delimited by Tabs, Spaces, and JSON data.

1. SEQUENCEFILE FORMAT

The data that is stored in sequence file format can be compressed and stored and hence usefull in minimizing the space taken by the files. Direct loading of the files is not possible as the files are stored in the binary. Hence the value in the sequence file can put using the insert command. Sequence files are flat files consisting of binary key-value pairs. When Hive converts queries to MapReduce jobs, it decides on the appropriate key-value pairs to be used for a given record. Sequence files are in the binary format which are able to split and the main use of these files is to club two or more smaller files and make them as a one sequence file.

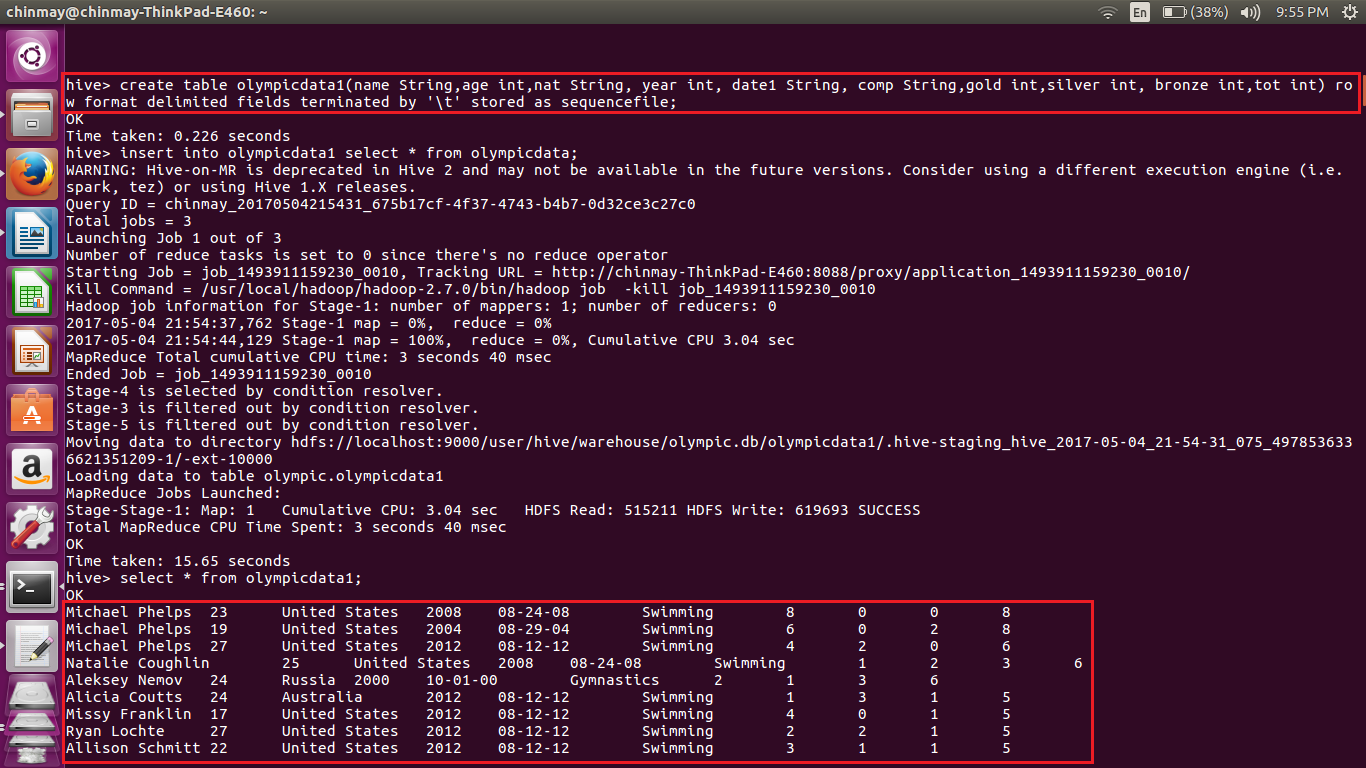
1. RCFILE FORMAT

RCFILE is used when we want to perform operations on multiple rows at a time. It stores the values in key value pairs. RCFILE stores columns of a table in form of record in a columnar manner. RCFILE first stores the metadata of a row split, as the key part of a record, and all the data of a row split as the value part. This means that RCFILE encourages column oriented storage rather than row oriented storage.

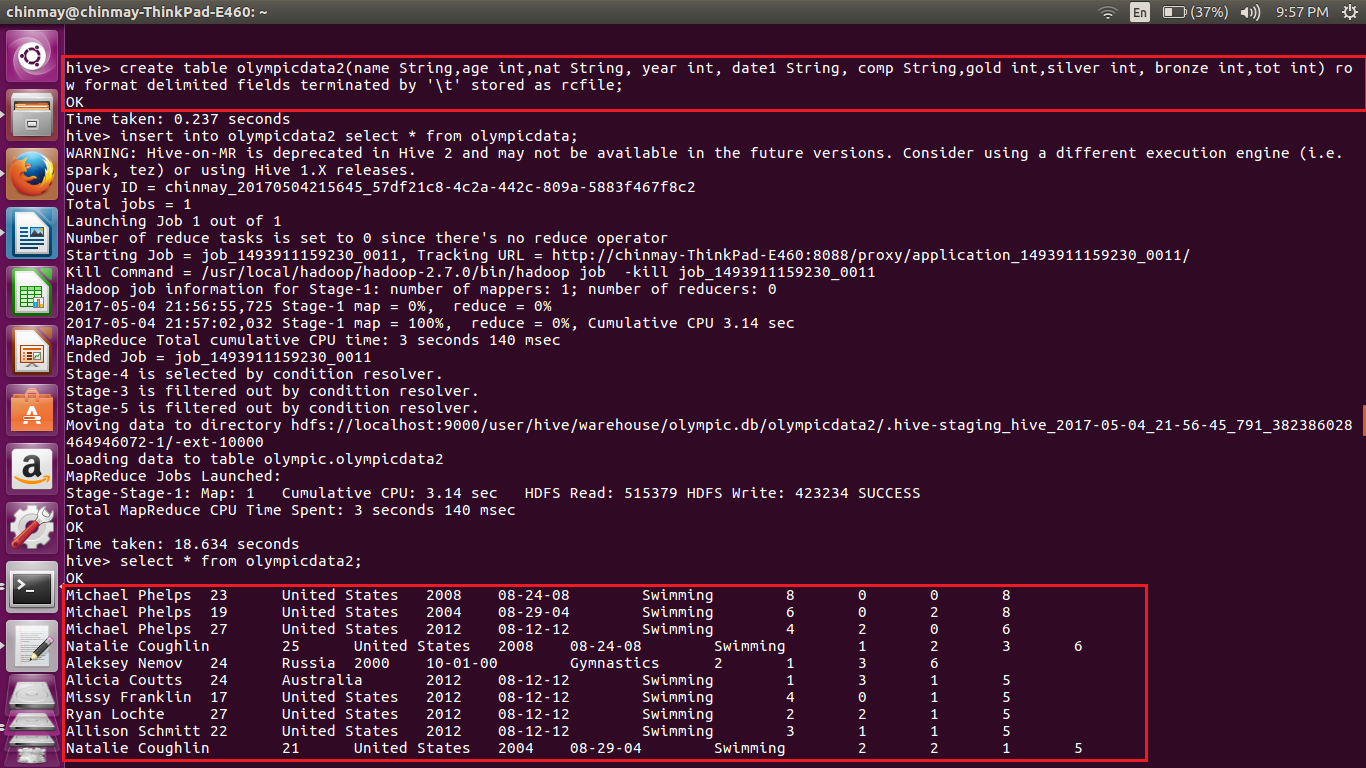
1. ORC FORMAT

ORC is similar to that of the RC file format but as the name says it is optimized one . it can store data in an optimized way than the other file formats. ORC reduces the size of the original data up to 75%.Because of this speed of the processing is also very fast. An ORC file contains rows data in groups called as Stripes along with a file footer. ORC format improves the performance when Hive is processing the data.

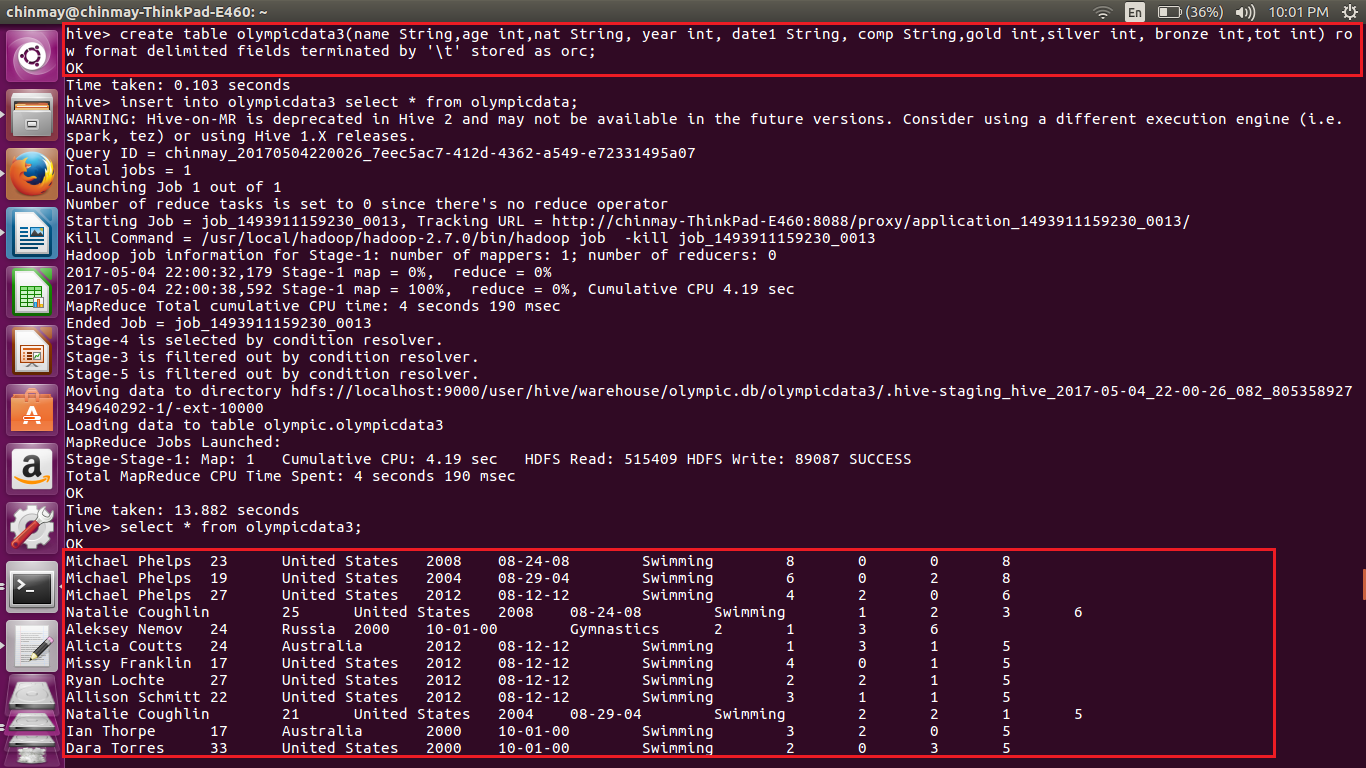
● SEQUECEFILE



● RCFILE



● ORC FILE



Share the commands used with their results.