**Note**: all the numeric values in equations below are presented as examples only.

= size of raw data

= size of processed data

= maximum size of data in hadoop

= size of temporary data

= size of intermediate data

= replications in hadoop

= available size required to process data of size in hadoop

= ***lzo*** compression ratio

= ***parquet*** compression ratio

(based on our numeric assumptions)

,

,