

Twitter faced an issue of an I/O performance limit, simply put, Twitter had issues to scale their systems, and simply scaling the system vertically was no help either as the bandwidth of HDD's are set to a stagnant number. The engineers tested out selectively placing the temp data in the YARN directory in a separate SSD. This gave incredible results, as now the temp MapReduce files and HDFS weren't contending for HDD space anymore, therefore, HDD utilization dropped significantly, and Hadoop could serve data faster. It also turned out, that the number of HDD's per cluster could be reduced with no penalty, with an SSD storing temp data from YARN. Furthermore, they implemented a way where, if the SSD was full, the data would get flushed to a safe location, preventing critical temp data being lost.