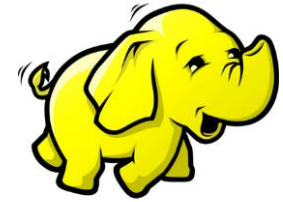


# Spark VS MapReduce



Speed	<ul style="list-style-type: none"><li>• 100x times faster when runs in memory</li><li>• 10x times faster when runs on disk</li></ul>	<ul style="list-style-type: none"><li>• It reads &amp; writes from the disk, hence it is slow compared to spark</li></ul>
Difficulty	<ul style="list-style-type: none"><li>• It provides high level operators such as map, filter, etc., and libraries.</li></ul>	<ul style="list-style-type: none"><li>• Need to hand code each operation in Hadoop</li></ul>
Processing	<ul style="list-style-type: none"><li>• It supports both batch and real-time processing</li></ul>	<ul style="list-style-type: none"><li>• It supports only batch processing</li></ul>
Handling Large Datasets	<ul style="list-style-type: none"><li>• Since it process data in memory, if data exceeds the available RAM then it causes problem</li></ul>	<ul style="list-style-type: none"><li>• Since it process data from disk, even if memory exceeds, it can read and write data from the disk. But the process is slow</li></ul>