**2. Explain the difference between MapReduce 1 and MapReduce 2 / Yarn:**

**In MapReduce1** (also called as Hadoop 1) where the HDFS Resource Management and Scheduling) and MapReduce are tightly couple. Because of this non-batch applications cannot be run on the Hadoop 1. It has single namenode so, it doesn’t provide High Availability and scalability.

**In MapReduce2** (also called Hadoop 2) where the resource management and scheduling tasks are separated from MapReduce which is separated by YARN and scheduling layers lies beneath the MapReduce layer. It also provides high system availability and scalability as we can create redundant NameNodes. The latest feature of snapshot through which can take backup of filesystems which helps disaster recovery.