The Google File System，一台电脑上的Block，在分布式系统中是一个Chunk， GFS由一个master和大量的chunkserver组成。在master中保留每个Chunk保存在哪个server中的位置信息。由chunkserver自行在内部进行检索，降低了master和server之间的耦合度，如果server中出现文件信息更改等操作，并不需要修改master中保存的目录索引。Master服务器主要负责为Tabletf服务器分配Tablets，检测新加入或者过期失效的Tablet服务器，对Tablet服务器进行负载均衡，通过轮询检测各个Tablet状态，并及时进行处理，为Tablet重新分配服务器，尽可能保证每个Tablet都有三份备份。

BigTable是基于GFS上的数据模型，内部按照字典顺序，对于Web的相关应用就可以保证域名相近的网页信息存储在附近的空间中，减少磁盘IO的时间，为了加快写数据的速度，每次写都在内存中的memtable, 写满了后直接写入磁盘成为一个SSTable，采用SSTable之间无序，但内部有序的方式进行组织，又为了避免这种组织方式导致的每次查询都需要查询所有的SSTable, 又加入了Bloom过滤器，只在“有可能”的SSTable中进行检索。即保证了插入数据时的效率，又保证了检索的速度。

MapReduce核心采用了分治的策略。将任务分解然后在多台处理能力较弱的计算节点中同时处理，然后将结果合并从而完成大数据处理。Map负责发任务，最后Reduce来将任务回收。