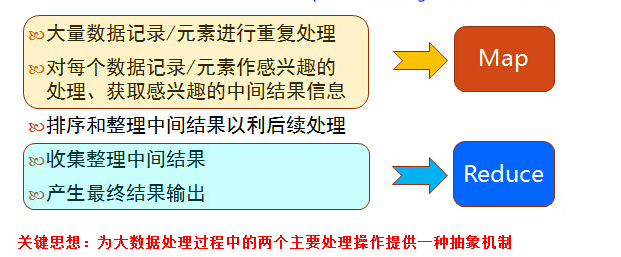
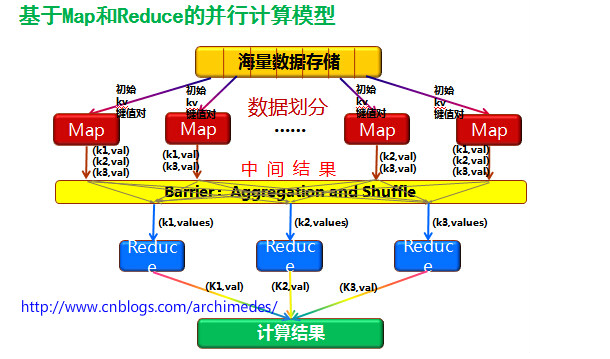
MapReduce主体思想：通过分散计算来分析大量数据。

一个大数据若可以分为具有同样计算过程的数据块，并且这些数据块之间不存在数据依赖关系，则提高处理速度的最好办法就是并行计算。





MapReduce常见算法：（https://www.cnblogs.com/1130136248wlxk/p/4975116.html）

1. 单词计数
2. 数据去重
3. 排序（按某个Key进行升序或降序排列）
4. 筛选（对源数据中所有数据进行排序，取出前K个数据）TOP K
5. 选择
6. 投影
7. 分组
8. 多表连接
9. 单表连接