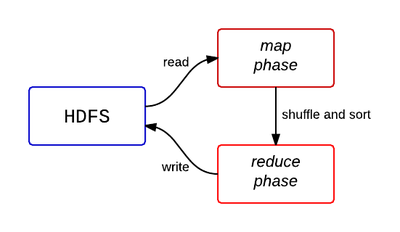
# Hadoop MapReduce

Veri kümeleri HDFS üzerinden yüklendikten sonra **Map**ve **Reduce**fazları işletilir. Bu kodlamaları Java, Pig ve Hive ve benzeri ile geliştirebiliriz.



Hadoop MapReduce ise HDFS üzerindeki büyük dosyaları verileri işleyebilmek amacıyla kullanılan bir yöntemdir. İstediğiniz verileri filtrelemek için kullanılan Map fonksiyonu ve bu verilerden sonuç elde etmenizi sağlayan Reduce fonksiyonlarından oluşan program yazıldıktan sonra Hadoop üzerinde çalıştırılır. Hadoop Map ve Reduce’lerden oluşan iş parçacıklarını küme üzerinde dağıtarak aynı anda işlenmesini ve bu işler sonucunda oluşan verilerin tekrar bir araya getirilmesinden sorumludur. Hadoop’un gücü işlenen dosyaların her zaman ilgili düğümün (node) yerel diskinden okunması ile ağ trafiğini meşgul etmemesinden ve birden fazla işi aynı anda işleyerek doğrusal olarak ölçeklenmesinden geliyor diyebiliriz. Yani aşağıdaki grafikte olduğu gibi Hadoop kümesindeki düğüm sayısı arttıkça performansı da doğrusal olarak artmaktadır.

MapReduce, **JobTracker** ve **TaskTracker** süreçlerinden oluşur. JobTracker yazılan MapReduce programının küme üzerinde dağıtılarak çalıştırılmasından sorumludur. Ayrıca dağıtılan iş parçacıklarının çalışması sırasında oluşabilecek herhangi bir problemde o iş parçacığının sonlandırılması ya da yeniden başlatılması da JobTracker’ın sorumluluğundadır. TaskTracker, DataNode’ların bulunduğu sunucularda çalışır ve JobTracker’dan tamamlanmak üzere iş parçacığı talep eder. JobTracker, NameNode’un yardımıyla DataNode’un local diskindeki veriye göre en uygun Map işini TaskTracker’a verir. Bu şekilde verilen iş parçacıkları tamamlanır ve sonuç çıktısı yine HDFS üzerinde bir dosya olarak yazılarak program sonlanır.

