当我看到Google的这三大论文时，我的内心是震撼的，Google竟奠定了风靡全球的大数据算法的基础。

第一篇讲的是GFS（谷歌分布式文件系统）。GFS是一个面向大规模数据密集型应用的分布式文件系统。GFS 虽然运行在廉价的硬件设备上，但是它依然了提供很好的可靠性，可以面对全球用户提供实时服务。它打破了单机存储形态，为了保证存储数据的安全，将数据分为多个部分，分别存储在数据块服务器的不同地方，当需要提取数据时，各个数据块服务器会按照目录将分块的数据重新进行拼凑，使多个系统高度并行，一起交给GFS的客户。那么GFS客户就可以通过直接索引文件名来获取所需要的数据。

第二篇论文讲的是Google Bigtable。Bigtable是一个分布式的多维度排列Map，被用来处理海量数据。论文描述了Bigtable提供的简单的数据模型，利用模型用户可以动态的控制数据的分布和格式。对于不同的应用需求，无论是数据量的批处理，还是响应速度上都可以很好的满足。它的设计目的是处理大量数据，并且具有很广泛的应用性，能够应用在上千台机器和多个产品上。

最后一篇讲是Google MapReduce。MapReduce是一个编程模型，也是一个处理和生成超大数据集的算法模型的相关实现。MapReduce编程模型的原理是：利用一个输入key/value pair集合来产生一个输出的key/value pair集合。创建一个Reduce函数用来合并所有的具有相同中间key值的中间value值。MapReduce在 Google 内部也已经成功应用于多个领域。

说起来Google确实是为大数据时代的发展做出了很大的贡献，毕竟Google的三个产品：Google File System、Map Reduce和Big Table的应用还是很广泛的。就比如GFS,这是一个可扩展的分布式文件系统，用于大型的、分布式的、对大量数据进行访问的应用。它运行于廉价的普通硬件上，通过软件的方式自动容错，它将服务器故障视为正常现象，通过软件的方式自动容错，在保证系统可靠性和可用性的同时，大大降低系统的成本。另一方面，它拥有着数据完整性，用于大型的、分布式的、对大量数据进行访问的应用。它又有着有效的诊断工具，广泛而细致的诊断日志以微小的代价换取了在问题隔离、诊断、性能分析方面起到了重大的作用。GFS服务器用日志来记录显著的事件，如服务器停机和启动和远程的应答。远程日志记录机器之间的请求和应答，通过收集不同机器上的日志记录，对它们进行分析恢复，就可以完整地重现活动的场景，并用此来进行错误分析。这些都给用户提供了总体性能较高的服务，比较方便快捷。

说到此，另两个产品Map Reduce和Big Table，它们则是基于GFS研发的。这三大基础核心技术构建出了完整的分布式运算架构。像Map Reduce，它是一种编程模型，用于大规模数据集的并行运算。Map（映射）和Reduce（归约），是它们的主要思想概念，这是从函数式编程语言借鉴的，并且还有矢量编程语言里的特性。它极大地方便了编程人员在不会分布式并行编程的情况下，将其程序运行在分布式系统上。 目前软件实现是指定一个Map函数，把一组键值对映射成一组新的键值对，指定并发的Reduce函数，用来保证所有映射的键值对中的每一个共享相同的键组。而Big Table则是分布式数据存储系统，用来处理海量的数据的一种非关系型的数据库。Big able是非关系型数据库，是一个稀疏的、分布式的和持久化存储的多维度排序Map。它适用于廉价设备，适合大规模海量数据以及分布式、并发数据处理，易于扩展，效率极高，支持动态伸缩。它们的功能确实极为强大，称它们奠定了大数据算法的基础一点都不为过。

通过对Google的这三大论文的阅读，我也是燃起了这方面的兴趣，尽管目前我对于它们的一些概念还是一知半解的，但这并不妨碍我了解它们，我也是在此浏览的过程中受益匪浅。