**关于三篇论文**

**19301029 陈旭**

论文1：The Google File System

主要讲述了一个可扩展的分布式文件系统——Google File System的开发原因、实现效果以及设计过程，重点关注了支持分布式应用程序的文件系统接口扩展。

在文中，说明了GFS的开发前景，然后从组件故障、数据量、新数据添加、增强灵活性四个方面讲述了开发GFS的必要性，依靠GFS可有效解决多客户端访问、磁盘空间等一系列问题。

关于GFS开发，作者先从对操作和功能的假设开始，逐步明细接口和体系结构等信息。其中我着重阅读了Consistency Model这一块，得知GFS有一个宽松的一致性模型，可以很好地支持高度分布式应用程序，且有着相对简单和高效的实现方式。而后文章讨论了GFS的保障以及它们对应用程序的意义，这部分中主要讲了数据突变的概念和GFS对于数据突变的应对方法以及GFS适应宽松的一致性模型的方法。

文章后续描述了关于GFS的交互、操作、容错判断以及测试几个方面。其中关于Micro-benchmarks，作者测试了多个客户端同时读/写文件，读入结果基本吻合，而写入结果比想象中速度要慢，但根据作者所说，几乎不影响实际使用。在构建和部署GFS的过程中，作者团队遭逢了关于操作或技术层面的各种各样的问题。比如磁盘与Linux直接相关，偶发的不匹配可能造成数据损坏。对此，开发团队通过校验和检测数据损坏情况，并修改内核以处理协议不匹配的问题。

综合本文，GFS已成功满足了开发团队的存储需求，并作为研发和生产数据处理的存储平台广泛应用于谷歌。 它是一个重要的工具，使用户能够在整个网络的规模上继续创新和解决问题。

论文2：MapReduce

主要讲述了一个编程模型——MapReduce，它是一个用于处理和生成大型数据集的关联的实现。这个项目的主要贡献是实现一个简单而强大的接口，可以实现大规模计算的自动并行化和分布，结合这个接口的实现，可以在大型集群的商用pc上实现高性能。

关于MapReduce的编程模型，可分为两个方面：Map和Reduce。Map由用户编写，它接受一个输入对，而pro得到一组中间的键值对。 MapReduce库将所有与中间键I关联的中间值组合在一起，并将它们传递给Reduce函数。 Reduce函数也是由用户编写的，它接受一个中间键I和该键的一组值。 它将这些值合并在一起，形成一个可能更小的值集。 通常，每次Reduce调用只产生0或1个输出值。 中间值通过iterator提供给用户的reduce函数。

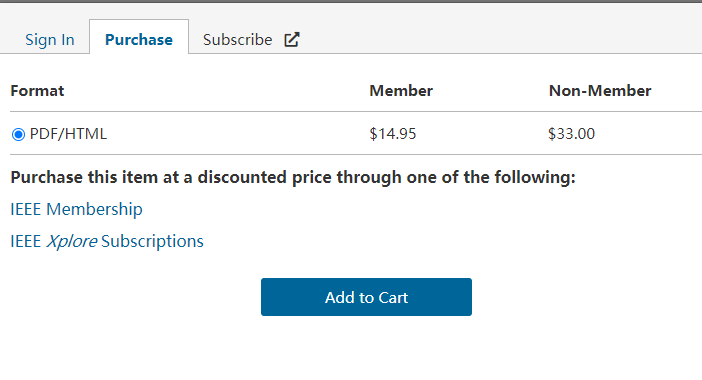
关于二者，作者设计了相关的类型，以规范其键和值的域。

总体工作流程为：通过自动划分数据，Map调用多台机器上的分布形成一组M分割，输入分块由不同的机器并行处理。用户程序中的MapReduce库首先将输入文件分成M个部分，然后在一个集群上启动程序的许多副本。而后由其中的主副本给其他副本指派工作，传输数据。

之后，作者从九个方面，讲述了关于Map和Reduce的一些拓展，并针对其性能等方面进行了测试。

在早期，开发团队就已经发现了MapReduce库对于当前计算机和计算领域能够处理的问题的适用范围是非常广阔的，它允许没有分布式和或并行系统经验的程序员也能够轻松利用大量资源。日前，它已经被广泛应用于各个领域。

论文3：FastSGG

资源要花钱，所以只了解了其摘要内容，抱歉

FastSGG是利用度分布生成模型的社交图生成器，用于完成各种社交网络分析任务，生成大规模的社交图。其生成方法主要包括两步：确定源顶点的出度和确定目标顶点以构造边。其中，提出了一种度分布生成（D2G）模型，用以加快图的生成过程。根据实验结果表明，FastSGG可以生成具有小世界性质、幂律度分布和社区结构的高质量社会图。

据我猜测，FastSGG的用途可能与QQ的好友推荐，WX的摇一摇有些类似，即根据用户的需求返回用户可能需要的结果。而社交图的应用理应更为宽泛，在图中，用节点或点表示个人，他们的连接由节点之间绘制的线表示，根据用户的设定和要求，用一张或多张图表来描述每个人的社交联系。社交图的使用，将使人们更容易与整个社交网络在线互动，由此，FastSGG的使用应该也是前途无量的。