谷歌三驾马车论文读后感

21世纪，数据就是石油，甚至比是由还珍贵！现如今，大数据已经应用在我们生活的方方面面了，它让我们的生活更加便捷，而且也为我们的生活质量提供了强有力的保障。就比如今年疫情期间，国家推出的健康宝，就是大数据的产物，它为我们国家的防疫工作做出了巨大的贡献。同时它也为商家挖掘潜在消费者提供了便利，马云曾说过，大数据让我知道了每个省女性的罩杯大小，由此可见大数据的可怕，因此需要我们对其进行限制，将其用在正道上。

其实我之前也稍微接触了一点大数据，那是在大二小学期的时候。当时我们做了一个天气预测的网站，我们组将从中国天气网上下载的数据输入到已经写好的算法中去，对未来的天气进行预测。结果让我大吃一惊，预测结果居然不超过3摄氏度，于是我第一次真切体会到了大数据的威力。

读完三篇论文后，我不禁感叹，谷歌作为一个科技公司，对于大数据的发展居然会有如此大的贡献。oogle的三篇重要论文分别是Google FS、MapReduce、BigTable，这三篇论文奠定了风靡全球的大数据算法的基础。

GFS发表于2003年，GFS系统由单个Master和大量块服务器构成，Master存放文件系统的所有元数据，而客户端从Master获取目标数据块的位置信息后，直接和块服务器交互进行读取操作。是一个可以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。它虽运行于价格低廉的普通硬件上，但却可以提供容错功能。这使部件的错误不再被当作异常，而是将其作为常见的情况加以处理。这使它可以给大量的用户提供总体性能较高的服务。

Mapreduce发表于2004年，它的模型分map、reduce两部分。它以GFS为基础，利用分布式计算思想，把大量数据拆分给多个低性能电脑，处理后的结果汇总输出。所以我们只需要输入需要定义任务、去除错误的数据，静等结果即可。由于参与运算的计算机都是性能较低的计算机，有时master会罢工，这时候所有的worker就会投票选出一个新的master。为防止worker罢工，就设置了一个程序让worker每隔一定时间向master发送信号，如果master没有接收到信号，master就会将任务分配给其他的worker。这可以使低性能的电脑省去昂贵的成本，用多台电脑的处理方式增加了容错率，并行处理大大减少了处理时间。

Bigtable发布于2006年，它是建立在 GFS 和 MapReduce之上的，是一个分布式的结构化的存储系统，是一个大型的具有容错性和自治性的系统。它可以可靠的处理PB级的数据和能够部署到千万台机器上。目前Google的很多项目都使用Bigtable存储数据。

读完三篇文章，我对大数据更加的向往，希望有一天我也可以从事相关的工作，看到自己可以随心所欲的操控数据，对各个方面进行分析和预测，别提多有成就感了！