21世纪，世界已经进入数据大爆炸的时代，大数据时代已经来临。从商业公司内部的各种管理和运营数据，到个人移动终端与消费电子产品的社会化数据，再到互联网产生的海量信息数据等，每天世界上产生的信息量正在飞速增长。2009年数据信息量达到8 000亿GB，而到2011年达到1.8 ZB。图灵奖获得者Jim Gray提出的“新摩尔定律”：“每18个月全球新增信息量是计算机有史以来全部信息量的总和”，已经得到验证。

大数据的“大”不仅仅体现在数据的海量性，还在于其数据类型的复杂性。随着报表、账单、影像、办公文档等在商业公司中得到普遍使用，互联网上视频、音乐、网络游戏不断发展，越来越多的非结构化数据进一步推动数字宇宙爆炸。数据海量而复杂，这是对大数据的诠释。

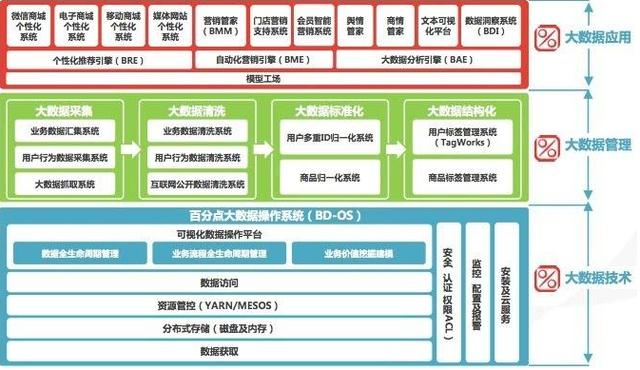
与传统的数据相比，大数据具有规模性（Volume）、多样性（Variety）、高速性（Velocity）和低价值密度（Value）的4V特点。规模性和高速性是数据处理一直以来研究和探讨的问题，多样性和价值密度低是当前数据处理发展中不断显现出来的问题，而且在可以预见的未来，随着智慧城市、智慧地球等各种新设想的不断成为现实，上面的4中问题将会变得更加凸显，而且是不得不面对的问题。

在这里我还是要推荐下我自己建的大数据学习交流群：前面是5322在加上181最后是47，群里都是学大数据开发的，如果你正在学习大数据 ，小编欢迎你加入，大家都是软件开发党，不定期分享干货（只有大数据开发相关的），包括我自己整理的一份2018最新的大数据进阶资料和高级开发教程，欢迎进阶中和进想深入大数据的小伙伴。

数据的产生经历了被动、主动和自动3个阶段。大数据的迅猛发展是信息时代数字设备计算能力和部署数量指数增长的必然结果。解决大数据研究中的问题，必须要从大数据的产生背景进行研究。大数据的产生源于规模效应，这种规模效应给数据的存储、管理以及数据的分析带来了极大的挑战，数据管理方式上的变革正在酝酿和发生。大数据的规模效应要求其存储、运算方案也应当从规模效应上进行考虑。传统的单纯依靠单设备处理能力纵向发展的技术早已经不能满足大数据存储和处理需求。以Google等为代表的一些大的数据处理公司通过横向的分布式文件存储、分布式数据处理和分布式的数据分析技术很好的解决了由于数据爆炸所产生的各种问题。

1 大数据关键技术

1.1 大数据系统的架构



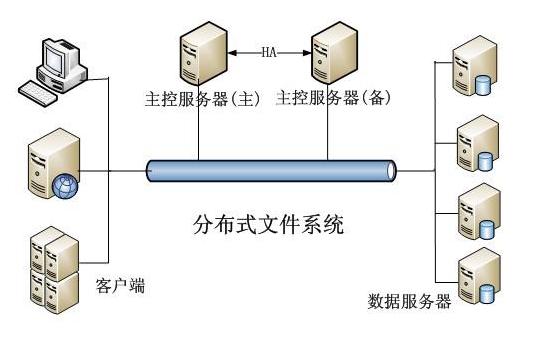
大数据处理系统不管结构如何复杂，采用的技术千差万别，但是总体上总可以分为以下的几个重要部分。

从数据处理的一般流程可以看到，在大数据环境下需要的关键技术主要针对海量数据的存储和海量数据的运算。传统的关系数据库经过近40年的发展已经成为了一门成熟同时仍在不断演进的数据管理和分析技术，结构化查询语言（SQL）作为存取关系数据库的语言得到了标准化，其功能和表达能力也得到的不断增强。但是，关系数据管理系统的扩展性在互联网环境下遇到了前所未有的障碍，不能胜任大数据分析的要求。关系数据管理模型追求的是高度的一致性和正确性。纵向扩展系统，通过增加或者更换CPU、内存、硬盘以扩展单个节点的能力，终会遇到“瓶颈”。

大数据的研究主要来源于依靠数据获取商业利益的大公司。Google公司作为全球最大的信息检索公司，其走在了大数据研究的前沿。面对呈现爆炸式增加的因特网信息，仅仅依靠提高服务器性能已经远远不能满足业务的需求。如果将各种大数据应用比作“汽车”，支撑起这些“汽车”运行的“高速公路”就是云计算。正是云计算技术在数据存储、管理与分析等方面的支持，才使得大数据有用武之地。

Google公司从横向进行扩展，通过采用廉价的计算机节点集群，改写软件，使之能够在集群上并行执行，解决海量数据的存储和检索功能。2006年Google首先提出云计算的概念。支撑Google公司各种大数据应用的关键正是其自行研发的一系列云计算技术和工具。Google公司大数据处理的三大关键技术为：Google文件系统GFS[4]、MapReduce[5]和Bigtable[6]。Google的技术方案为其他的公司提供了一个很好的参考方案，各大公司纷纷提出了自己的大数据处理平台，采用的技术也都大同小异。下面将从支持大数据系统所需要的分布式文件系统、分布式数据处理技术、分布式数据库系统和开源的大数据系统Hadoop等方面介绍大数据系统的关键技术。

1.2 分布式文件系统



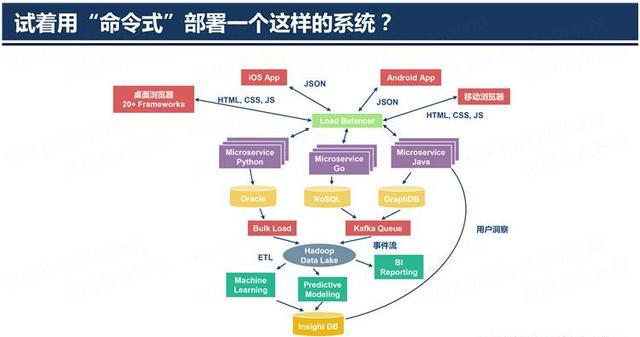
文件系统是支持大数据应用的基础。Google是有史以来唯一需要处理如此海量数据的大公司。对于Google而言，现有的方案已经难以满足其如此大的数据量的存储，为此Google提出了一种分布式的文件管理系统——GFS。

GFS与传统的分布式文件系统有很多相同的目标，比如，性能、可伸缩性、可靠性以及可用性。但是，GFS的成功之处在于其与传统文件系统的不同。GFS的设计思路主要基于以下的假设：对于系统而言，组件失败是一种常态而不是异常。

GFS是构建于大量廉价的服务器之上的可扩展的分布式文件系统，采用主从结构。通过数据分块、追加更新等方式实现了海量数据的高效存储，所示给出了GFS体系结构。但是随着业务量的进一步变化，GFS逐渐无法适应需求。Google对GFS进行了设计，实现了Colosuss系统，该系统能够很好地解决GFS单点故障和海量小文件存储的问题。

除了Google的GFS，众多的企业和学者也从不同的方面对满足大数据存储需求的文件系统进行了详细的研究。微软开发的Cosmos支撑其搜索、广告业务。HDFS、FastDFS、OpenAFS和CloudStore都是类似GFS的开源实现。类GFS的分布式文件系统主要针对大文件而设计，但是在图片存储等应用场景中，文件系统主要存储海量小文件，Facebook为此推出了专门针对海量小文件的文件系统Haystack。

通过多个逻辑文件共享同一个物理文件，增加缓存层、部分元数据加载到内存等方式有效地解决了海量小文件存储的问题。Lustre是一种大规模、安全可靠的，具备高可靠性的集群文件系统，由SUN公司开发和维护。该项目主要的目的就是开发下一代的集群文件系统，可以支持超过10 000个节点，数以拍字节的数量存储系统。



1.3 分布式数据处理系统

大数据的处理模式分为流处理和批处理两种。流处理是直接处理，批处理采用先存储再处理。

流处理将数据视为流，源源不断的数据形成数据流。当新的数据到来即立即处理并返回所需的结果。大数据的实时处理是一个极具挑战性的工作，数据具有大规模、持续到达的特点。因此，如果要求实时的处理大数据，必然要求采用分布式的方式，在这种情况下，除了应该考虑分布式系统的一致性问题，还将涉及到分布式系统网络时延的影响，这都增加了大数据流处理的复杂性。目前比较有代表性的开源流处理系统主要有：Twitter的Storm、Yahoo的S4以及Linkedin的Kafka等。

Google公司2004年提出的MapReduce编程模型是最具代表性的批处理模型。MapReduce架构的程序能够在大量的普通配置的计算机上实现并行化处理。这个系统在运行时只关心如何分割输入数据，在大量计算机组成的集群上的调度，集群中计算机的错误处理，管理集群中的计算机之间必要的通信。

对于有些计算，由于输入数据量的巨大，想要在可接受的时间内完成运算，只有将这些计算分布在成百上千的主机上。这种计算模式对于如何处理并行计算、如何分发数据、如何处理错误需要大规模的代码处理，使得原本简单的运算变得难以处理。MapReduce就是针对上述问题的一种新的设计模型。

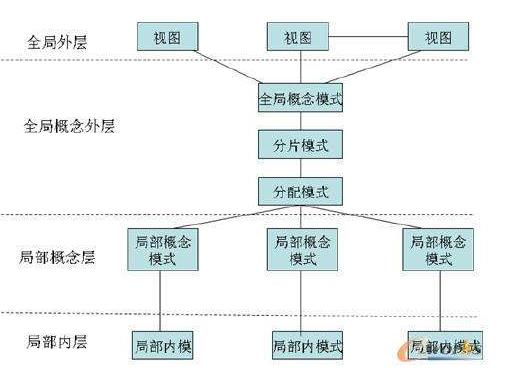
MapReduce模型的主要贡献就是通过简单的接口来实现自动的并行化和大规模的分布式计算，通过使用MapReduce模型接口实现在大量普通的PC上的高性能计算。

MapReduce编程模型的原理：利用一个输入键-值（Key/Value）对集合来产生一个输出的key/value对集合。MapReduce库的用户用两个函数表达这个计算：Map和Reduce。用户自定义的Map函数接受一个输入的key/value值，然后产生一个中间key/value对集合。MapReduce库把所有具有相同中间key值的value值集合在一起传递给Reduce函数。用户自定义的Reduce函数接收一个中间key的值和相关的一个value值的集合。Reduce函数合并这些value值，形成一个较小的value值集合

MapReduce的提出曾经遭到过一系列的指责和诟病。数据专家Stonebraker就认为MapReduce是一个巨大的倒退，指出其存取没有优化、依靠蛮力进行数据处理等问题。但是随着MapReduce在应用上的不断成功，以其为代表的大数据处理技术还是得到了广泛的关注。研究人员也针对MapReduce进行了深入的研究，目前针对MapReduce性能提升研究主要有以下几个方面：多核硬件与GPU上的性能提高；索引技术与连接技术的优化；调度技术优化等。在MapReduce的易用性的研究上，研究人员正在研究更为高层的、表达能力更强的语言和系统，包括Yahoo的Pig、Microsoft的LINQ、Hive等。

除了Google的MapReduce，Yunhong Gu等人设计实现了Sector and Sphere云计算平台[18]，包括Sector和Sphere两部分。Sector是部署在广域网的分布式系统，Sphere是建立在Sector上的计算服务。Sphere是以Sector为基础构建的计算云，提供大规模数据的分布式处理。Sphere的基本数据处理模型如图4所示。

针对不同的应用会有不同的数据，Sphere统一地将它们以数据流的形式输入。为了便于大规模地并行计算，首先需要对数据进行分割，分割后的数据交给SPE执行。SPE是Sphere处理引擎，是Sphere的基本运算单元。除了进行数据处理外SPE还能起到负载平衡的作用，因为一般情况下数据量远大于SPE数量，当前负载较重的SPE能继续处理的数据就较少，反之则较多，如此就实现了系统的负载平衡。



1.4 分布式数据库系统

传统的关系模型分布式数据库难以适应大数据时代的要求，主要的原因有以下几点：

（1）规模效应带来的压力。大数据时代的数据远远超出单机处理能力，分布式技术是必然的选择。传统的数据库倾向于采用纵向扩展的方式，这种方式下性能的增加远低于数据的增加速度。大数据采用数据库系统应该是横向发展的，这种方式具有更好的扩展性。

（2）数据类型的多样性和低价值密度性。传统的数据库适合结构清晰，有明确应用目的的数据，数据的价值密度相对较高。在大数据时代数据的存在的形式是多样的，各种半结构化、非结构化的数据是大数据的重要组成部分。如何利用如此多样、海量的低价值密度的数据是大数据时代数据库面临的重要挑战之一。

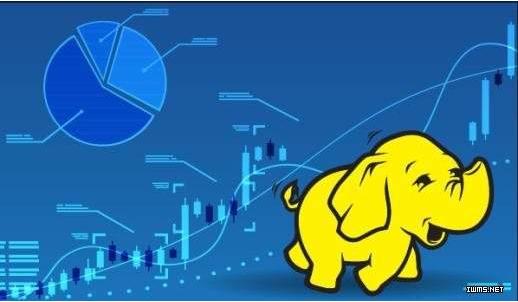
（3）设计理念的冲突。关系数据库追求的是“一种尺寸适用所有”，但在大数据时代不同的应用领域在数据理性、数据处理方式以及数据处理时间的要求上千差万别。实际处理中，不可能存在一种统一的数据存储方式适应所有场景。

面对这些挑战，Google公司提出了Bigtable的解决方案。Bigtable的设计目的是可靠的处理拍字节级别的数据，并且能够部署到千台机器上。Bigtable已经实现了以下几个目标：适用性广泛、可扩展、高性能和高可靠性。Bigtable已经在超过60个Google的产品和项目上得到了应用。这些产品在性能要求和集群的配置上都提出了迥异的需求，Bigtable都能够很好地满足。Bigtable不支持完整的关系数据模型，为用户提供了简单的数据模型，利用这个模型，客户可以动态控制数据的分布和格式。

用户也可以自己推测底层存储数据的位置相关性。数据的下标是行和列的名字，名字可以是任意的字符串。Bigtable将存储的数据都视字符串，但是Bigtable本身不去解释这些字符串，客户程序通常会把各种结构化或者半结构化的数据串行化到这些字符串。通过仔细选择数据的模式，客户可以控制数据的位置的相关性。最后，可以通过Bigtable的模式参数来控制数据是存放在内存中、还是硬盘上。Bigtable数据模型如图5所示，给出了Bigtable存储大量网页信息的实例。

除了Google公司为人熟知的Bigtable，其他的大型Internet内容提供商也纷纷提出大数据系统。具有代表性的系统有Amazon的Dynamo[19]和Yahoo的PNUTS[20]。Dynamo综合使用了键/值存储、改进的分布式哈希表（DHT）、向量时钟等技术实现了一个完全的分布式、去中性化的高可用系统。PNUTS是一个分布式的数据库系统，在设计上使用弱一致性来达到高可用性的目标，主要的服务对象是相对较小的记录，比如在线的大量单个记录或者小范围记录集合的读和写访问，不适合存储大文件、流媒体。

Bigtable、Dynamo、PNUTS等技术的成功促使研究人员开始对关系数据库进行反思，产生了一批为采用关系模型的数据库，这些方案通称为：NoSQL（not only SQL）。NoSQL数据库具有以下的特征：模式只有、支持简易备份、简单的应用程序接口、一致性、支持海量数据。目前典型的非关系型数据库主要有以下集中类别。



1.5 大数据系统的开源实现平台

Hadoop

除了商业化的大数据处理方案，还有一些开源的项目也在积极的加入到大数据的研究当中。Hadoop[22]是一个开源分布式计算平台，它是MapReduce计算机模型的载体。借助于Hadoop，软件开发者可以轻松地编出分布式并行程序，从而在计算机集群上完成海量数据的计算。Intel公司给出了一种Hadoop的开源实现方案，如图6所示。

在该系统中HDFS是与GFS类似的分布式文件系统，它可以构建从几台到几千台常规服务器组成的集群，并提供高聚合输入输出的文件读写访问。HBase[23]是与Bigtable类似的分布式、按列存储的、多维表结构的实时分布式数据库。可以提供大数据量结构化和非结构化数据的高度读写操作。Hive[24]是基于Hadoop的大数据分布式数据仓库引擎。它可以将数据存放在分布式文件系统或分布式数据库中，并使用SQL语言进行海量信息的统计、查询和分析操作。ZooKeeper[25]是针对大型分布式系统的可靠协调系统，提供的功能包括：配置维护、名字服务、分布式同步、组服务等。

它可以维护系统配置、群组用户和命名等信息。Sqoop[26]提供高效在Hadoop和结构化数据源之间双向传送数据的连接器组件。它将数据传输任务转换为分布式Map任务实现，在传输过程中还可以实现数据转换等功能。Flume[27]是分布式、高可靠的和高可用的日志采集系统，它用来从不同源的系统中采集、汇总和搬移大量日志数据到一个集中式的数据存储中。

主要结合大数据的产生背景、需求和系统结构，介绍了当前全球在大数据技术方面的进展情况。从分析可以看到，大数据系统的解决方案必将落地于现有的云计算平台。云计算平台的分布式文件系统、分布式运算模式和分布式数据库管理技术都为解决大数据问题提供了思路和现成的平台。

通过分析也可以看到，大数据的问题的研究，必然是以商业利益为驱动，一些大的依靠数据牟利的大公司必然会是大数据应用的主体，大数据一定会成为重点研究领域。

总的来说，目前对于大数据的研究仍处于一个非常初步的阶段，还有很多问题需要解决，有兴趣的朋友，可以关注我们的多智时代，及时查阅人工智能和大数据的小知识。