关于云计算的海量数据存储模型

0 引言

随着越来越多的人使用计算机，整个网络会产生数量巨大的数据，如何存储网络中产生的这些海量数据，已经是一个摆在面前亟待解决的问题。现在常见的三种存储方式是DAS[1]、NAS 和SAN，但是面对网络产生的越来越多的数据，这三种方式的缺点就明显的暴露出来。DAS 存储方式可扩展性差，系统性能低，存储分散。NAS 虽然使用方便，成本低廉，但最是存储性能差。SAN 存储效能优异，能大幅提升网络上工作效能与资料传输效率，但是其架构为封闭式架构，无法整合不同系统，且规模过大成本较高。

2006 年底，Google 第一次提出了“云[2]”的概念，为我们更好的处理网络中产生的海量数据带来了希望。

本文提出的基于云计算的海量数据存储模型，是依据云计算的核心计算模式MapReduce[3]，并依托实现了MapReduce 计算模式的开源分布式并行编程框架Hadoop[3]，将存储模型和云计算结合在一起，实现海量数据的分布式存储[4]。

1 云计算

云计算[5]是一种计算模式，也是一种全新的商业模式。云计算（Cloud Computing）是分布式处理（Distributed Computing）、并行处理（Parallel Computing）和网格计算（GridComputing）的发展或者说是这些计算机科学概念的商业实现。

云计算[6]是随着网络中产生的越来越多的数据而被提出的，在云计算中，无数的软件和服务都置于云中，这里的云是指可以自我维护和管理的虚拟计算资源。这些软件和服务均构筑于各种标准和协议之上，可以通过各种设备来获得。

云计算是一种超级的计算模式，可以把网络中的计算机虚拟为一个资源池，将所有的计算资源集中起来，并用特定软件实现自动管理，使得各种计算资源可以协同工作，这就使得处理数量巨大的数据成为了可能。

2 一级标题基于云计算的海量数据的存储

2.1 MapReduce 模式

MapReduce 是云计算的核心计算模式，是一种分布式运算技术，也是简化的分布式编程模式，用于解决问题的程序开发模型，也是开发人员拆解问题的方法。

MapReduce 模式的主要思想是将自动分割要执行的问题（例如程序），拆解成Map（映射）和Reduce（化简）的方式。MapReduce 的流程所示：

在数据被分割后通过Map 函数的程序将数据映射成不同的区块，分配给计算机机群处理达到分布式运算的效果，在通过Reduce 函数的程序将结果汇整，从而输出开发者需要的结果。

MapReduce 借鉴了函数式程序设计语言的设计思想，其软件实现是指定一个Map 函数，把键值对(key/value)映射成新的键值对(key/value)，形成一系列中间结果形式的key/value 对，然后把它们传给Reduce(规约)函数，把具有相同中间形式key 的value 合并在一起。Map 和Reduce 函数具有一定的关联性。函数描述所示：

2.2 Hadoop 框架

Hadoop 是一个实现了MapReduce 计算模型的开源分布式并行编程框架，程序员可以借助Hadoop 编写程序，将所编写的程序运行于计算机机群上，从而实现对海量数据的处理。

此外，Hadoop 还提供一个分布式文件系统(HDFS）及分布式数据库（HBase）用来将数据存储或部署到各个计算节点上。Hadoop 框架如所示：

借助Hadoop 框架及云计算核心技术MapReduce 来实现数据的计算和存储，并且将HDFS 分布式文件系统和HBase 分布式数据库很好的融入到云计算框架中，从而实现云计算的分布式、并行计算和存储，并且得以实现很好的处理大规模数据的能力。

2.3 基于云计算的海量数据存储模型

根据数据的海量特性，结合云计算技术，特提出基于云计算的海量数据存储模型，如所示在中，主服务控制机群相当于控制器部分，主要负责接收应用请求并且根据请求类型进行应答。存储节点机群相当于存储器部分，是由庞大的磁盘阵列系统或是具有海量数据存储能力的机群系统，主要功能是处理数据资源的存取。HDFS 和Hbase 用来将数据存储或部署到各个计算节点上。Hadoop 中有一个作为主控的Master，用于调度和管理其它的计算机（将其称之为TaskTracker），Master 可以运行于机群中任一台计算机上。TaskTracker 负责执行任务，必须运行于DataNode 上，DataNode 既是数据存储节点，也是计算节点。Master将Map 任务和Reduce 任务分发给空闲的TaskTracker，让这些任务并行运行，并负责监控任务的运行情况。如果其中任意一个TaskTracker 出故障了，Master 会将其负责的任务转交给另一个空闲的TaskTracker 重新运行。用户不直接通过Hadoop 架构读取及HDFS 和Hbase存取数据，从而避免了大量读取操作可能造成的系统拥塞。用户从Hadoop 架构传给主服务控制机群的信息后，直接和存储节点进行交互进行读取操作。

2.4 数据存取算法基本思想

存数据算法基本思想为:

1 存储数据时，将存储数据的信息及其附加信息（如用户ID）发送给主服务控制机群。

2 主服务控制机群接收到数据的信息。

3 将接收到的数据信息传送给Hadoop 架构。

4 MapReduce 利用其Map 函数对数据进行切块计算。

5 HDFS 和Hbase 根据节点状态将数据均衡分配到各存储节点。

6 将数据块信息及存储节点地址返回主服务控制机群，并由主服务控制机群反馈给用户。

7 用户为每个存储节点建立一个数据块队列，将数据块并行上传到对应的存储节点。

因为 Hadoop 具有高容错性，能自动处理失败节点，所以当发现某个节点失效时，立即将正在上传的部分数据块进行重新分配。

取数据算法基本思想为:

1 下载文件时，将要下载的文件信息传送给主服务控制机群。

2 主服务控制机群接收到要下载的文件信息。

3 HDFS 和Hbase 查找该文件的块信息，并且将查找到的信息反馈给主服务控制机群。

4 主服务控制机群然后把信息传回给用户。

5 用户根据接收到的主服务控制机群传回的信息，为每个存储节点创建一个下载线程，将文件块并行下载到本地计算机临时文件夹中。

6 用户在下载完所有文件块以后，根据MapReduce 的Reduce 函数整合成一个完整的文件，并删除文件块。

当 Hadoop 发现某个节点失效时，立即将正在下载的文件交由另一空闲的节点来重新进行下载，从而保证下载顺利完成。

Hadoop 具有高容错性，能自动处理失效节点是通过MapReduce 来实现的。MapReduce通过把对数据集的大规模操作分发给网络上的每个节点实现可靠性，每个节点会周期性的把完成的工作和状态的更新报告回来。如果一个节点保持沉默超过一个预设的时间间隔，主节点记录下这个节点状态为死亡，并把分配给这个节点的数据发到别的节点。此外每个操作要保证不会发生并行线程间的冲突。

3 结果与分析

本课题是由南京市卫生局牵头，涵盖的市县二甲、三甲医院达到十几个，涉及的数据包括医院电子病历系统，HIS，PACS 系统，数据量巨大。利用各个医院的硬件资源搭建一个Hadoop 的平台，整个平台由各个医院的服务器系统和汇聚到卫生局信息中心的交换机构成，使用的操作系统

为linux redhat fedora，Java 环境为jdk-1_5_0-linux，Hadoop 软件版本为hadoop-0.19.1。其中Hadoop 的主要配置文件hadoop-site.conf 配置如下：

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl"href="configuration.xsl"?>
<configuration>
<property>
<name>fs.default.name</name>
<value>hdfs://10.40.33.11/</value>
</property>
<property>
<name>mapred.job.tracker</name>
<value>hdfs://10.40.33.12/</value>
</property>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
```

以上配置文件只是配置了Hadoop 的HDFS 中Namenode 的位置、MapReduce 中的tasktraker 的位置以及备份数量。

与云计算系统相比，云存储可以认为是配置了大容量存储空间的一个云计算系统。从架构模型来看，云存储系统比云计算系统多了一个存储层，同时，在基础管理也多了很多与数据管理和数据安全有关的功能，两者在访问层和应用接口层则是完全相同的。

4 结论

本文给出了很少一部分医院的医疗数据，如何扩大到全市所有的医院，还有待进一步的研究。

总体上讲，云计算领域的研究还处于起步阶段，尚缺乏统一明确的研究框架体系， 还存在大量未明晰和有待解决的问题，研究机会、意义和价值非常明显。现有的研究大多集中于云体系结构、云存储、云数据管理、虚拟化、云安全、编程模型等技术，但云计算领域尚存在大量的开放性问题有待进一步研究和探索。

在职硕士论文

[参考文献] (References)

[1] 余顺争，谢长生．融合NAS 和SAN 的存储网络设计与实现[J].电子学报，2006，34(11)：2012-2017.

[2] 林立宇,陈云海,张敏等.云计算技术及运营可行性分析[J]. 广东通信技术,2008 (12): 33-38

[3] Dean J. MapReduce: Simplified Data Processing on Large Clusters[C]//Proc. of the 6th IEEE Symposium onOperatingSystem Design and Implementation. San Francisco, CA, USA:[s. n.], 2004.

[4]（美）Abraham Silberschatz．数据库系统概念[M]．机械工业出版社，2000.5-8

[5] 张健.云计算概念和影响力解析[J] .电信网技术,2009,(1):15-18.

[6] Rajkumar Buyya, CheeShin Yeo, Srikumar Venugopal. Market-Oriented Cloud Computing: Vision, Hype, andReality for Delivering IT

Services as Computing Utilities[C]//Proceedings of the 10th IEEE InternationalConference on High Performance Computing and Communications, HPCC2008:5-13, Dalian, China, Sept.2008