ISSN 1009-3044
Computer Knowledge and Technology 电脑知识与技术
Vol.7, No.1, January 2011,pp.16-19

E-mail: jslt@cccc.net.cn
http://www.dnzs.net.cn
Tel:+86-551-5690963 5690964

基于私有云的数据冗余技术研究

周 游,刘鹏,杨盛祥,薛志强,文艾

(解放军理工大学 军事网格研究中心,江苏 南京 210007)

摘要:作为云计算技术的一个分支,云存储的应用逐渐兴起。文章介绍了私有云存储的概念和特点,在此基础上结合私有云存储的环境,对几种数据冗余方式进行了分析比较。

关键词:云存储;数据冗余;纠删码

中图分类号:TP393 文献标识码:A 文章编号:1009-3044(2011)01-0016-04

Data Redundancy Strategy in Private Cloud Storage

ZHOU You, LIU Peng, YANG Sheng-xiang, XUE Zhi-qiang, WEN Ai

(MilGrid Research Center, PLA Univ. of Sci. & Tech. , Nanjing 210007, China)

Abstract: A branch of cloud computing called cloud storage was rising, this paper introduced the concept and characteristics of private cloud storage. Several data redundancy strategies are discussed and compared in private cloud storage environment.

Key words: cloud storage; data redundancy; erasure codes

云存储是云计算运用的一个分支,突出了存储的特点,是云计算技术在存储领域的运用,它是通过集群管理、网格技术、并行处理和分布式处理,将网络中大量相连的各类存储设备综合集成起来协同工作,向用户提供一个虚拟化的存储空间服务和业务访问功能。由于在网络上充斥的海量数据处理,使得以数据存储和管理为核心的云存储技术愈来愈受到重视,以 Google、EMC、亚马逊、微软等为代表的众多知名 IT 企业和科研机构相继开展了研究。

数据存储服务是云存储服务商向用户提供的重要服务内容,因此数据的可靠性是用户非常关心的一个性能,数据冗余技术因此成为云存储以致整个云计算运用中的关键技术。选用好的冗余方案,对于系统的鲁棒性、资料的完整性、处理数据的吞吐性能都有影响,本文将对云存储常用的数据冗余技术进行分析,结合私有云的特性,讨论私有云的数据冗余策略。

1 私有云存储简介

1.1 私有云存储的概念

私有云建立在相关单位的防火墙后面,采用单位自身的软硬件设施,可以部署在企业数据中心或相似地点,通过内部管理人员或专业云存储服务商提供技术支持来进行统一控制,对存储空间进行集中分配,来满足内部人员不同权限和功能的需求。相比于公共云,私有云可谓"麻雀虽小、五脏俱全",而其更好地满足了用户心理上的隐私安全感,内部用户容易对其进行了解掌握,也被军队等对权限限制要求较高的机构所青睐。

企业级云运用的大量普及才是云存储的前景,要想让企业真正信服云计算的价值、就必须推广私有云存储技术。

1.2 私有云存储系统的特点

与传统的存储系统相比,私有云存储有其特点,而这些正是其吸引用户的独有优势:

1) 易于扩充,成本低廉:私有云存储采用的是并行扩容技术,其容量分配不受物理硬盘限制,可以很方便地扩充容量和性能,对存储设备升级不会导致服务中断,这样可以减少各单位存在的硬盘空间浪费,用户根据自己的需要向服务器弹性地申请所需要的空间,降低了用户的使用成本。对于用户来说,这是云存储吸引的一个重要因素。

2) 方便管理,可靠性高:易于管理是云存储系统设计时重点考虑的问题,相比分散零落的小集群,云存储系统后面有着专业的维护人员,使得用户可以减少维护数据时间。同时,数据采用集中存储的方式,由数据中心的管理员对数据进行统一管理、负责资源的分配、负载的均衡、软件的部署、安全的控制,并能更可靠地进行数据安全的实时监测以及数据的及时备份和恢复,降低了数据被盗、被破坏和外泄的可能,降低控管风险。

3) 分布于网络,移动方便:私有云存储系统以网络为依托,用户一旦将自己的资料存入云中,只要能连上网络,便可以在随处、任意的存取自己数据,既方便了用户,也减少了用户携带移动存储设备带来的设备丢失损坏等隐患[1]。对于军队来说,在战场环境中,只要能接入战术互联网,就有可能上传下载信息,提高了效率和战场生存能力。

4) 灵活多变的用户群:相比公共云,私有云在集群数量和网络带宽上都有更大的弹性,从用几十台机器搭建小规模云的小公司到有成千上万台主机的大公司,都可以是私有云存储服务的使用者。

收稿日期:2010-10-28

作者简介:周游(1984-),男,江西宜春人,硕士,主要研究方向为分布式存储;刘鹏(1970-),男,四川绵阳人,博士,教授;杨盛祥(1984-),男,安徽合肥人,硕士;薛志强(1985-),男,河南洛阳人,硕士;文艾(1981-),男,湖南长沙人,硕士,讲师。

2 私有云存储中的数据冗余方式

总体而言,云存储服务都采用了分布式数据存储方式,通过将冗余数据分散存储在连入云系统的大量存储设备中去,以提高系统的抗摧毁性,目前主要有以下冗余方案:

2.1 完全副本备份冗余

完全副本备份冗余方式通过生成多个数据块的拷贝,将之存储在不同的存储服务器上,以达到备份的目的,只要有一个副本保持完整,数据块就可以正常获取。对于完全副本来说,主要涉及到的技术有副本生成数目的选取和分布策略的制定。

很多著名的存储系统如 GFS、Atmos、ceph 都采用了完全副本冗余方式,在 Google 公司的 GFS[2]中,通过完全复制数据块,在云存储系统中为其生成三个副本;EMC 的 Atmos 则为付费用户提供比免费用户更多的副本服务;Amazon 的 S3[3]系统对每个用户数据产生多个副本,来保证用户数据的安全性。采用完全副本冗余的方式,有以下几个优势:

1) 负载均衡:在面对访问频率较高的请求时,由于存在多个副本,可以有效地降低单个节点出现的负载过重的瓶颈现象,使系统负载均衡。

2) 降低访问延迟:元数据服务器在收到客户的访问请求后,可以就近分配访问延迟最小的存储副本节点给客户,提高了性能,而且可以节省网络带宽。

3) 处理简单:完全副本冗余只是将数据块进行了最简单的完全复制,相比其他的编码,减轻了处理器的负担,在系统实现上也较为简单。

完全副本备份技术简单直观,比较适合高负荷的系统,但是其对存储空间的消耗大,带来的是存储节点和电能消耗的增多;且随着系统冗余度的提高,其性能呈现明显下降,同时由于私有云中用户的带宽和存储空间有限,无限制地增加副本数量,将使得网络开销增大,网络性能明显降低。若采用惯用的三个副本备份方式,冗余度 n:k=3,在一个 PB 级别的系统中,仅有三分之一的空间作为使用,造成了浪费,这对于小公司来说,性价比太低。尽管实现简单,在最新版的 HDFS 系统中仍然考虑不单纯依赖完全副本冗余方式来进行数据冗余[4]。

2.2 纠删码冗余

纠删码(erasure codes)具备识别错码和纠正错码的功能,当错码超过纠正范围时可把无法纠错的信息删除。纠删码将一份数据分解成 m 块,通过编码将其转化成 n 块,只需要 t(t≥m)块数据存在,便可以恢复出原始数据。

可以说任何一种纠错码都具有纠删码的功能,但是,如何设计一种纠删能力强、编译速度快、冗余度低、满足实时高速率传输的纠删码以应用到云存储系统中去,是一个难题。下面将介绍几种常见的纠删码,其中,部分已经在存储系统中有所运用。

2.2.1 几种常见的纠删码编码

RS(Reed-solomon)[5]编码是一种经典的纠错码,由 Reed 和 Solomon 在 1960 年提出,它是一种前向纠错算法,能够从收到的 t 个数据块中恢复原有的数据,主要用于数据通信和存储中。RS 编码具有可靠性高、任意冗余度、空间最优等特点,并且实现简单。

编解码性能较低是 RS 编码的缺点,由于所有运算都是在伽罗华域中展开,在进行乘除运算时需要进行两次伽罗华域转换,使得计算复杂,而不适合对实存储系时性能要求高的统使用。

根据生成矩阵的不同,RS 编码可以分为 Vandermonde 编码和 Cauchy 编码:

1) 基于范德蒙德矩阵的 RS 编码(VRS)

若选取编码生成矩阵为 Bk×n,使得 BT=(bi,j),其中 bi,j=(bi)j-1,bi∈B(pr)(p 为素数,r 为正整数),则称如下所示的矩阵为范德蒙德矩阵。

$$B=\begin{bmatrix} b_0^0 & b_1^0 & b_2^0 & \cdots & b_{n-1}^0 \\ b_0^1 & b_1^1 & b_2^1 & \cdots & b_{n-1}^1 \\ b_0^2 & b_1^2 & b_2^2 & \cdots & b_{n-1}^2 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ b_0^{k-1} & b_1^{k-1} & b_2^{k-1} & \cdots & b_{n-1}^{k-1} \end{bmatrix}$$

在有限域 F 上,设 Ik×k 为单位矩阵,Bk×(n-k)为范德蒙德矩阵,若取生成矩阵 G =(I|G),则所得纠删码为范德蒙德码[6]。VRS 编码算法有如下特点:使用范德蒙德矩阵来计算与维护校验码;在 GF 中进行数学运算;使用高斯消元从错误中恢复。

2) 基于柯西矩阵的 RS 编码(CRS)

设$\{x_1,x_2,\cdots,x_m\}$和$\{y_1,y_2,\cdots,y_n\}$是有限域 F 中两个元素集,若对(1)$\forall i,j\in\{1,2,\cdots,m\}$,有 $x_i+y_j\neq 0$ (2) $\forall i,j\in\{1,2,\cdots,m\}(i\neq j)$有 $x_i\neq y_j$ 和 $\forall i,j\in\{1,2,\cdots,n\}(i\neq j)$有 $y_i\neq y_j$,则称如下图所示的矩阵为域 F 上的柯西矩阵。

$$C=\begin{bmatrix} \frac{1}{x_1+y_1} & \frac{1}{x_1+y_2} & \cdots & \frac{1}{x_1+y_n} \\ \frac{1}{x_2+y_1} & \frac{1}{x_2+y_2} & \cdots & \frac{1}{x_2+y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{x_m+y_1} & \frac{1}{x_m+y_2} & \cdots & \frac{1}{x_m+y_n} \end{bmatrix}$$

在有限域 F 上,设 Ik×k 为单位矩阵,Gk×(n-k)为柯西矩阵,若取生成矩阵 G =(I|G),则所得纠删码为柯西码。CRS[7]在 VRS 基础上有两个改进,一是使用柯西矩阵代替范德蒙德矩阵,二是把 GF 域上的运算全部转换为异或运算。柯西编码极大地提高了 RS 编码的

性能。

LDPC 码[9](Low-Density Parity-Check)是一种线性分组编码,其校验矩阵 0 多 1 少,呈现稀疏矩阵特征,因此称为低密度校验码。

Tornado[9]编码是一种速度非常快的 LDPC 编码,它基于稀疏矩阵,对数据对象及与其相近的数据对象进行异或操作生成检验码,并且通过不规则级联二分图对校验码进行编码,使得可靠性得到保证。在编译码过程中主要进行了异或操作,时间开销与块的数量成线性增长,因此可以较快地恢复原有对象。由于 Tornado 校验矩阵的稀疏性,其编解码性能呈线性增长,具有比 RS 编解码更优良的性能[10],但由于 Digital Fountain 对算法版权的保护,有关资料较少。

由于 Tornado 编码的特点,相比 RS 编码,若与待恢复字符相关的几个字符被摧毁,其可能面临满足冗余度的情况下,却无法恢复出数据,虽然其可以通过多级联方式来保证可靠性,但也因此,Tornado 码被证明必须要获得 k (1 +ε)个节点才能完成解码,相比较,RS 只需要 k 个。其局限性[11]在于为了获取更快的速度,必须收到略大于原文件的数据,这样,在同可靠性情况下,冗余度比 RS 高。

2.3 RAID 冗余

在分布式存储系统中,RAID(Redundant Array of Inexpensive Disks)技术运用的比较广泛,早期使用最简单的奇偶校验码保证数据的可靠性,从 RAID0-RAID6,技术愈加成熟,RAID5 在 N+1 的情况下,允许出现 1 块磁盘损坏,而 RAID6 中则允许出现 2 块磁盘损坏。

但是随着磁盘阵列不断增加,多个磁盘同时失效的几率越来越高,使得存储系统可靠性能迅速下降。在最近,RAID 技术又被加以创新,新版的 HDFS 中就将重复副本冗余技术和 RAID 技术结合起来,在 google 下一代的 GFS 中,已经考虑采用 raid 和纠删码技术来减少系统的空间消耗。

3 几种冗余方式的比较

3.1 可靠性分析

完全副本冗余将存储数据的完整副本存放在各自独立的节点上,只要一个节点有效,就可以满足系统的可用性,设系统的可靠性为 R,节点的可靠度都为 u,每个对象数据的副本拷贝数量为 k,假设各个节点的可靠性都是独立的,只有所有存储节点都坏掉数据才会损坏,所以系统的可靠性为:

$$R=1-(1-u)^k$$

若要达到的系统冗余度 R 是个定值,由上述计算求得所需要的副本数 K 值为:

$$k=\log\frac{1-R}{1-u}$$

在纠删码的冗余方案中,将对象分成 m 个,通过编码将其转换为 n 个数据块,冗余度 k=n/m,只要任意 t(t≥m)个数据块完整便可以恢复出原始对象,因此,系统的可靠性为:

$$R=\sum_{i=m}^{km}C_{km}^{i}u^{i}(1-u)^{km-i}$$

代入 K 值得:

$$R=\sum_{i=m}^{n}C_{n}^{i}u^{i}(1-u)^{n-i}$$

在 RAID 中,以 RAID6 为例,其在 RAID5 的基础上把校验信息由一位增加到两位,要求至少 4 块硬盘,允许坏掉两块,因此,设需要冗余的硬盘数为 n 块,它的冗余度为 k=(n+2)/n,设每块硬盘的可靠性能为 u,由于 RADI6 最多允许坏两块硬盘,则系统的可靠性为:

R=1- P{2 个磁盘以上失效}= P{0 个磁盘失效} +P{1 个磁盘失效} + P{2 个磁盘失效},即:

$$R=\sum_{i=0}^{2}C_{n+2}^{i}u^{n+2-i}(1-u)^{i}\quad(n\geq 2)$$

当 n 的值越大时,RAID 的 R 值呈明显下降,因此一个 RAID 的 n 值不易过大。

3.2 性能对比

很多研究人员对分布式存储系统的冗余方式进行过研究,但由于系统动态环境参数的不确定性以及采用的不同分析方法,对纠删码和完全副本两种冗余方式的结论并不完全一致:

Weatherspoon[12]在 2002 年通过量化的方法对完全副本和纠删码两种冗余方法进行了比较,发现若要使分布式系统达到相同的可靠性,纠删码方案比完全副本使用空间少,并且维护系统占用的带宽较小,由此得出结论:同冗余度下,纠删码冗余方案的可靠性较高。

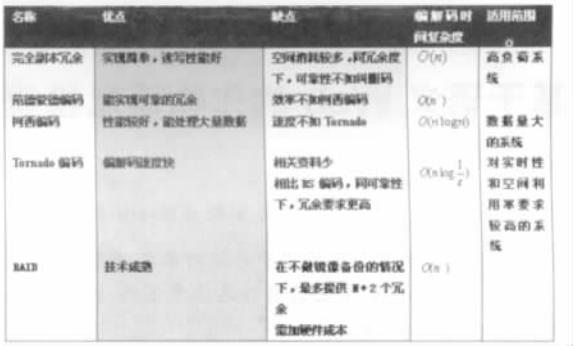
而 Rodrigues[13]等人则指出,只有在节点平均可用性水平较低时,纠删码方案才优于完全副本,反之,使用纠删码方案反而增加了系统设计的复杂性、占用了更多的网络带宽。

总体来说,完全副本方案实现简单,在高负荷系统中具有良好的响应性能;而纠删码方案有利于提高系统空间的使用率,并且提供了更高的可靠性,但是可能带来系统设计的复杂性;RAID 有广泛运用的历史,但不能满足新的存储需求,具体选择何种冗余方案,应该根据系统的实际情况确定。

3.3 综合比较

综合比较上述几种冗余方案,其优缺点见右图。

对于私有云存储用户来讲,在保证可靠性的同时,追求高效的空间利用率和较强的抗毁能力是两个主要关注的焦点,因此,采用纠删码技术能满足以上两点需求,关键就是如何找到一种性能更快,更可靠的编解码算法,柯西编码和 Tornado 算法所展现出来的特点值得进一步研究。



名称	优点	缺点	编解码时间复杂度	适用范围
完全副本冗余	实现简单,读写性能好	空间消耗较多,同冗余度下,可靠性不如纠删码	$O(n)$	高负荷系统
范德蒙德编码	能实现可靠的冗余	效率不如柯西编码	$O(n)$	
柯西编码	性能较好,能处理大量数据	速度不如 Tornado	$O(n\log n)$	数据量大的系统
Tornado 编码	编解码速度快	相关资料少 相比 RS 编码,同可靠性下,冗余要求更高	$O(n\ln\frac{1}{\varepsilon})$	对实时性和空间利用率要求较高的系统
RAID	技术成熟	在不做镜像备份的情况下,最多提供 $n+2$ 个冗余 需加硬件成本	$O(n)$	

4 结束语

目前,如何管理好日益增长的海量资料已经成为各个公司顺利展开业务的关键点,各公司针对自身情况,纷纷提出了对存储系统新的要求。在这种情况下,本文对目前主流的数据冗余方式进行了分析比较,尤其针对私有云存储的特点,提出了思路。

参考文献:

[1] 陈全,邓倩妮. 云计算及其关键技术[J].计算机应用,2009,29(9).

[2] Sanjay Ghemawat,Howard Gobioff,Shun-Tak Leung.The Google file system[M].Proc.of the 19th ACM SOSP.New York:ACM Press,2003:29-43.

[3] Amazon.Amazon Simple Storage Service (S3)[EB/OL].http://www.amazon.com/s3,2010.

[4] WICKER SB,BHARGAVA VK.Reed-Solomon Codes and Their Applications[M].New York:IEEE Press,1994.

[5] Rizzo L.On the Feasibility of Software FEC[Z].Internal report University of Pisa,1997.

[6] Blomer J,Mitzenmacher M,Shokrollahi A.An XOR-based erasure-resilient coding scheme,ICSI Technical Report,No.TR-95048[R],1995.

[7] Gallager R G.Low-Density Parity-Check Codes[M].MIT Press,Cambridge,MA,1963.

[8] Byers W,Luby M,Mitzenmacher M,et,al.A Digital Fountain Approach to Reliable Distribution of Bulk Data[Z].SIGCOMM,1998:56-67.

[9] 慕建君,王鹏,王新梅.正则低密度纠删码的性能分析[J].西安电子科技大学学报,2003(40):469-472.

[10] 孙伟平,汤毅凡.基于 Tornado 码的存储冗余算法研究[J].微处理机,2008(2):71-74.

(上接第 8 页)

10)$L_k=\{c \in C_k \mid c.count \geq min\_sup\}$;

11)end;

12)return $L=\cup_k L_k$; //L 为输出结果

其中 apriori_gen 实现连接和剪枝两个动作,由 L_{k-1} 得到 C_k。在连接部分产生可能的候选;在剪枝部分使用 Apriori 性质删除具有非频繁子集的候选。最后经过多次迭代得到对任务发布者和威客都具有价值的高频数据项集。

关联规则发现任务的本质是要在数据库中发现强关联规则。应用在威客网站中关联规则的发现,就是要找到威客对网站上各种任务文件之间访问的相互联系。例如,用关联规则发现技术,我们可以找到以下的相关性:40%的威客访问页面/ employer1/task1 时,也访问了/ employer1/ mission2。30%的客户在访问/ employer2/job1 时,在/ employer2/task1 进行了在线任务接收。那么,employer1 与 employer2 所发布的任务 task1、mission2 和 job1 必然有其相关性,比如都是属于同一种类型的、相似难度的或者具有某种内在的不容易被人们发现但却能够反应威客一定需求特征的关联规则[5]。利用这些规则,威客网站就可以对所有的具有相同或相似信息特征的威客有针对性地推荐最适合他们的任务,这样便于威客选择和完成最胜任的任务,提高威客的任务成功率和信心,避免浪费大量的时间和精力浏览和选择任务,而选择的任务也并非就是最适合自己的,同时,对于威客网站可以合理分配资源,更好的组织站点,提供有针对性的服务,实施有效的市场策略。

3 结束语

本文针对目前比较流行威客网站所存在的制约威客模式发展的问题,运用数据挖掘的技术,引入关联规则算法,通过找出威客和任务之间所存在的一些关联性,有效解决了威客网站的信息不对称和资源浪费等问题,实现威客网站资源的合理分配和有效利用。基于需求的威客网站 Web 数据挖掘应用很好地解决了从威客网站 Web 数据到有价值的知识转化的问题,为我国威客网站提供有针对性的个性化服务提供了技术上的可行性,并为威客网站提高资源利用率,更好地服务用户,实现经济效益增长提供了一条发展的新路。

参考文献:

[1] 毛国君.数据挖掘原理与算法[M].北京:清华大学出版社,2005.

[2] 张冬青.数据挖掘在威客网站中应用问题研究[J].现代情报.2005(9).

[3] 李凤慧.面向威客网站的 Web 数据挖据的研究[D].青岛:山东科技大学,2004.

[4] 陆垂伟.威客网站中数据挖掘技术的研究与应用[J].商场现代,2006(4).

[5] CHEN Yu-ru,HUNG Ming-chuan,Don-lin YANG.Using data mining to construct an intelligent web search system[J].International Journal of Computer Processing of Oriental Languages,2003,16(2).