

引用格式:李海荣,方中纯.一种改善云存储综合性能的体系结构及协调算法[J].微电子学与计算机,2020,37(5):23-27. [LI H R, FANG Z C. An architecture and coordination algorithms for improving the comprehensive performance of cloud storage [J]. Microelectronics & Computer, 2020, 37(5): 23-27.]

一种改善云存储综合性能的体系结构及协调算法

李海荣^{1,2}, 方中纯²

(1 北京科技大学 计算机与通信工程学院, 北京 100080;

2 内蒙古科技大学 工程训练中心, 内蒙古 包头 014000)

摘 要: 在深入研究云存储体系结构和影响系统整体性能相关因素的前提下, 提出一种基于多个主从结构的云存储系统而构建的一种分层控制体系结构的云存储系统. 然后以两种分布式文件系统构建的分层云存储系统为研究对象, 给出基于该体系结构的分布式文件系统的选择算法的数学模型以及相关公式, 应用案例对选择结果进行分析、测试. 最后, 从多个角度比较了 FastDFS、HDFS 和本文提出的分层次体系结构的性能, 实验结果证明该体系结构在云存储系统中综合性能比单个系统表现更好.

关键词: 云存储; 体系结构; 分布式文件系统; 协调

中图分类号: TP393

文献标识码: A

文章编号: 1000—7180(2020)05—0023—05

An architecture and coordination algorithms for improving the comprehensive performance of cloud storage

LI Hai-rong^{1,2}, FANG Zhong-chun²

(1 School of Computer & Communication Engineering, University of Science & Technology, Beijing 100080, China; 2 Engineering Training Center, Inner Mongolia University of Science & Technology, Baotou 014000, China)

Abstract: On the premise of in-depth study of cloud storage architecture and related factors affecting the overall performance of the system, a hierarchical control architecture of cloud storage system based on muti-master-slave structure is proposed by using the theory of Collaborative Artificial Intelligence. Secondly, taking two kinds of distributed file systems as the research object, the mathematical model and related formulas of the selection algorithm of distributed file system based on the architecture are given, and the selection results are analyzed and tested with a case. Finally, the performances of FastDFS, HDFS and the hierarchical architecture proposed in this paper are compared from various perspectives. The experimental results show that the overall performance of the architecture in cloud storage is better than that of a single system.

Key words: cloud storage; architecture; distributed file system; coordination

1 引言

随着信息技术与互联网的发展, 数据呈现爆炸

式地增长, 由此带来的存储压力, 越来越受到学术界和工业界的广泛关注. 而云存储(存储即服务)由于具有良好的可扩展性、对用户的透明性、按需分配的

收稿日期: 2019—09—10; 修回日期: 2019—10—15

基金项目: 国家自然科学基金(61462069); 内蒙古自然科学基金(2017MS0604; 2017MS(LH)0603); 内蒙古科技大学教改重点项目(JY2016003; JY2018013)

灵活性等优点,一跃成为云计算发展最迅猛的分支,成为带动公有云服务发展的核心动力,也代表着未来存储服务的发展趋势^[1].

目前较为典型和重要的云存储架构包括主从体系结构^[2],如 Google 公司的 GFS(Google File System)、HDFS(Hadoop Distributed File System)、FastDFS 等,以及分散体系结构(也称为对等体系结构),如亚马逊提出的 Dynamo.

在对多种分布式文件系统深入研究的基础上、利用协同人工智能的相关理论,提出了一种分层式、高性能的云存储结构(HPCS2A—High Performance Cloud Storage System Architecture)^[3-4],并基于该结构实现了协调算法,使系统整体性能得到提升.并对其进行实验验证和测试.

2 常见的云存储体系结构

2.1 主从体系结构

主从体系结构^[5]是分布式系统普遍采用的一种体系结构,主从体系结构的云存储系统如图 1 所示.

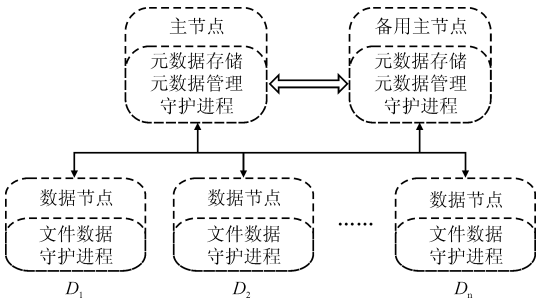


图 1 主从体系结构的云存储系统

在图 1 中主要包括三类节点:主节点、数据节点和备用主节点,它们在云存储系统中的主要功能如下:

(1)主节点,系统中只有一个主节点,该节点中保存数据的元数据,例如:用户信息、权限、复制因子、文件名、数据分块信息等.

(2)数据节点,这种节点用于保存数据对象的数据块,在系统中存在多个数据节点,用于确保系统的可靠性、高效率,一个数据块会分布在这些数据节点上.

(3)备用主节点,该节点是主节点的一个热备份,在系统中也是唯一的一个节点,目的是为了防止云存储系统发生单点故障.

主从式架构包含三类进程,它们在云存储系统中的主要功能如下:

(1)主节点的守护进程,主要用于监视云存储系

统中的文件变化、分配数据保存的数据节点、管理元数据、监视数据节点的运行状态等.

(2)数据节点的守护进程,主要用于定时发送心跳信息给主节点上报工作状态、根据主节点分配的任务接收客户端上传的数据、根据主节点分配的任务发送数据到客户端、维护该节点的数据块等.

(3)备用主节点的守护进程,在主节点正常工作的情况下,该进程负责从主节点拉取元数据到该备用节点上;如果在主节点失败的时候,该节点自动充当主节点的工作,完成主节点的任务,并且发出警告提示系统管理员处理故障,但是整个云存储系统仍然正常工作.采用这种热备份主节点的方式能够有效防止云存储系统的单点故障发生.

2.2 分散体系结构

在分散体系结构的云存储系统^[6-7]中,每个数据节点都是对等的,没有主从之分.如图 2 所示.

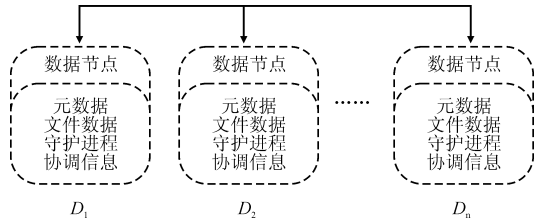


图 2 分散结构的云存储系统

每个数据节点都保存一份本节点数据块的元数据,同时这些数据节点之间进行两两通讯、相互作用、相互协调,每个节点都保存协调过程中的协调信息.协调信息中包含本节点收到的其它节点的数据(节点主机信息、心跳状态信息、元数据信息等).主要工作过程描述如下:

(1)从云存储中获取数据,在这种体系结构中获取数据可以从任意一个节点开始,一般选择距离客户端访问代价最小的节点.在该节点上查询元数据,如果找到对应数据块,那么直接获取数据,守护进程把本节点的数据块推送到客户端;反之,如果元数据中没有找到,就在协调信息中查找元数据,如果找到就告诉客户端到对应数据节点上获取数据块;如果寻找元数据失败,那么从协调信息中获取云存储系统的节点信息,重新选择一个节点再次开始前面的过程.

(2)保存数据到云存储系统中,如同读数据一样,可以任意选择一个节点开始.把数据块写入该节点中并追加元数据记录,数据块写入成功后,接着从协调信息中获取正常工作的节点,然后把该数据块

的数据推送到另一个数据节点中建立数据块副本。

(3)副本的放置方法,当数据块需要建立副本的时候,把数据块计算 Hash 值,然后根据 Hash 值从协调信息中选择工作节点,然后把副本放置到对应工作节点上。

3 HPCS2A 的体系结构

3.1 HPCS2A 的设计

本文设计的 HPCS2A 的一个重要特点就是整个系统是由 $m(m \geq 2)$ 个分布式存储系统构成的一个云存储系统^[8-10]。这种结构带来的优势是可以获取 $m(m \geq 2)$ 个分布式存储系统的最佳性能,从而给用户提供更好的存储和计算服务。由于系统使用了 m 个分布式存储系统,所以在整个系统的顶层增加了一个协调控制层,从而形成一个逻辑上单一的云存储系统。其架构如图 3 所示。

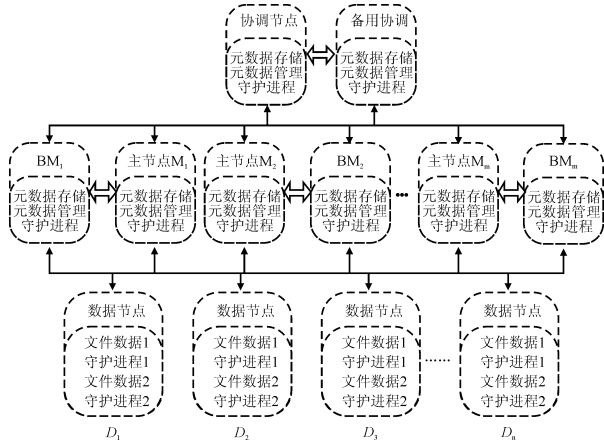


图3 分层控制体系结构的云存储系统 HPCS2A

本文把云存储的这种体系结构分为三层,从上到下依次为:

(1)第一层,该层是云存储系统的协调控制层,主要用于协调下层的 m 个分布式存储系统,并给用户提供一个逻辑唯一的云存储系统。同时,该层决定下层的分布式存储系统的选择方法,给出选择向量。选择向量的好坏决定着系统的整体性能。

(2)第二层,该层是分布式文件系统的主节点层,这一层由 m 对主节点 M 和备用主节点 BM 构成,每对主节点作为一个分布式文件系统的主控节点。通过上层的协调选择主节点,从而选择相应的分布式文件系统进行数据存储或者读取。

(3)第三层,该层是分布式文件系统的数据节点层,这一层用于保存文件的数据分块或者直接保存文件。同时,这些数据节点也是各个分布式文件系统

的共用数据节点。

3.2 HPCS2A 的工作过程

当有数据上传到系统中或者动态调整系统中的数据时,需要利用选择向量通过第一层即协调层来选择下层的分布式系统。选择向量的计算公式如式(1)所示:

$$\begin{aligned} cv &= f(dt \times DFSP) \\ &= f\left((c \ s) \times \begin{bmatrix} c_1 & c_2 & \cdots & c_m \\ s_1 & s_2 & \cdots & s_m \end{bmatrix}\right) \\ &= f((\gamma_1 \ \gamma_2 \ \cdots \ \gamma_m)) \end{aligned} \quad (1)$$

式中, cv 为选择向量(choice vector),协调层根据该向量决定选择下层哪个分布式文件系统; dt 为数据的分类向量 $(c \ s)$, c 表示计算性能需求, s 表示存储性能需求,二者和为 1; $DFSP$ 为分布式文件系统性能矩阵 $\begin{bmatrix} c_1 & c_2 & \cdots & c_m \\ s_1 & s_2 & \cdots & s_m \end{bmatrix}$, 这个矩阵是 m 个分布式文件系统的性能矩阵,其中每一列代表一个分布式文件系统的性能向量, c 代表分布式文件系统表现出来的计算能力优势, s 代表该分布式文件系统表现出来的存储性能优势,二者和为 1,如果 $c > s$,说明该节点计算能力强,反之,则说明该节点存储能力强。

f 为该函数是选择向量区分度函数,定义如下:

$$f \rightarrow \frac{\alpha}{1 + \exp(0.5 - x)} \quad (2)$$

式中, α 为区分度放大系数,其取值范围为 $[1, 100]$,通过该系数的调整可以得到区分度更好的选择向量。 x 为是式(1)计算得到的 $(\gamma_1 \ \gamma_2 \ \cdots \ \gamma_m)$ 向量。

所以,由式(1)和式(2)得如下公式:

$$\begin{aligned} cv &= \frac{\alpha}{1 + \exp\left[0.5 - \left((c \ s) \times \begin{bmatrix} c_1 & c_2 & \cdots & c_m \\ s_1 & s_2 & \cdots & s_m \end{bmatrix}\right)\right]} \\ &= \frac{\alpha}{1 + \exp(0.5 - (\gamma_1 \ \gamma_2 \ \cdots \ \gamma_m))} \\ &= (cv_1 \ cv_2 \ \cdots \ cv_m) \end{aligned} \quad (3)$$

协调层通过式(3)可以计算出分布式文件系统的选择向量 $(cv_1 \ cv_2 \ \cdots \ cv_m)$, 其中的 $cv_i (1 \leq i \leq m)$ 为对应分布式文件系统的选择因子,其取值范围为 $(0, 1)$,一般选择前 3 个最大的因子的分布式文件系统或者选择 $cv_i (cv_i \geq 0.5)$ 作为保存数据的分布式文件系统。

考虑到实际应用和复杂性,本文设定 m 为 2,并

选择一个存储性能表现较优的FastDFS和计算性能较优的 HDFS 组成一个云存储系统作为研究对象,那么协调层计算选择向量的公式如下:

$$cv = \frac{\alpha}{1 + \exp\left[0.5 - \left((c \ s) \times \begin{pmatrix} c_1 & c_2 \\ s_1 & s_2 \end{pmatrix}\right)\right]} \tag{4}$$

给出一个实例:假设有一个数据对象,它的分类向量为(0.7 0.3),云存储系统的性能矩阵为

$$\begin{pmatrix} 0.9 & 0 \\ 0.1 & 1 \end{pmatrix}, \text{那么}$$
$$cv = \frac{\alpha}{1 + \exp\left[0.5 - \left((c \ s) \times \begin{pmatrix} c_1 & c_2 \\ s_1 & s_2 \end{pmatrix}\right)\right]}$$
$$= \frac{30}{1 + \exp\left(0.5 - \left((0.7 \ 0.3) \times \begin{pmatrix} 0.9 & 0 \\ 0.1 & 1 \end{pmatrix}\right)\right)}$$
$$= (0.992 \ 0.002) \tag{5}$$

由式(5)计算的选择向量,由于 0.992≥0.5,所以协调层选择第 1 个分布式文件系统保存数据.

下面给出这个选择结果的分析步骤:
(1)分类向量为(0.7 0.3),因为 0.7≥0.3,说明这个数据是偏向于计算服务需求的.

(2)第 1 个分布式文件系统的性能向量为(0.9 0.1),由于 0.9≥0.1,说明该分布式文件系统在计算性能方面表现得比较优秀.

(3)式(5)的计算结果说明数据需要保存到第 1 个分布式文件系统中,即:计算服务需求大的数据放到计算性能较优的分布式文件系统,这个结果非常合理.

3.3 性能分析

HPCS2A 不是一种分布式文件系统,而是由多种分布式文件系统 DFS 共同组成的大系统,并结合这些 DFS 的优点展示出综合性能.通过对多种 DFS 进行协调和优选,使得存储数据的访问和计算都能达到相当高的效率.以下是对 HPCS2A 进行测试并与 FastDFS、HDFS 进行比较的情况.

测试环境:2 个主节点、9 个从节点.硬件配置: I5-6200 CPU 2.3GHz,16 GB 内存,CentOS64 位.测试项目:文件上传、文件下载、文件删除、计算能力.测试文件的个数与大小:分为三种情况,见表 1.系统中部署 FastDFS、HDFS 与 HPCS2A 协调系统,其中 HPCS2A 协调控制系统能够方便地开启或者关闭 FastDFS 和 HDFS,这为性能的测试带来了

便利.

表 1 测试结果表

项目	测试文件个数及大小	FastDFS /秒	HDFS /秒	HPCS2A /秒
上传文件测试	个数:2 500 个 大小:[64B,20MB]	38.98	271.98	40.37
	个数:100 个 大小:[500MB,2GB]	387.58	413.95	390.28
	个数:10 个 大小:2GB 以上	87.9	114.5	89.5
文件下载测试	个数:2 500 个 大小:[64B,20MB]	24.36	260.16	31.34
	个数:100 个 大小:[500MB,2GB]	240.96	461.77	245.86
	个数:10 个 大小:2GB 以上	92.14	128.21	93.28
删除文件	个数:2 500 个 大小:[64B,20MB]	1.24	210.34	211.32
	个数:100 个 大小:[500MB,2GB]	1	68.57	68.62
	个数:10 个 大小:2GB 以上	0.17	7.06	7.1
计算能力(对数据进行分布式异或运算)	个数:2 500 个 大小:[64B,20MB]	无框架,编程实现难,分析效率如同单机.		
	个数:100 个 大小:[500MB,2GB]		589.56	578.28
	个数:10 个 大小:2GB 以上		196.56	201.34

对测试结果进行分析,可以得到以下结论:
(1)FastDFS 主要针对数据存储的应用场景而产生的,它对小文件的存储效率是相当高的.
(2)HDFS 主要针对大数据处理而产生的,当文件比较大的时候 HDFS 也逐渐体现出效率的提高.
(3)在数据处理能力上,HDFS 自然显示出优点.
(4)使用 HPCS2A 协调控制后,整个系统会根据数据的使用情况自动选择性能优秀的 DFS.由于协调和选择会影响一点效率,当文件数量多并且文件小的时候,协调选择带来的性能影响相对增大,但是遇到批量的大文件,这点影响完全可以忽略.此外,由于有了 HPCS2A 的协调选择,整个大系统仍然表现出和 HDFS 一样的数据分析效率.
(5)使用 HPCS2A 后,删除效率降低,因为可能存在需要在多个 DFS 中进行数据的删除,但是这不影响系统的整体性能,因为云存储环境中可以采用惰性删除,即:先高速做出删除标志,事后再真实删除即可.

通过上述分析可知,当采用 HPSCS2A 进行多个 DFS 协调构成高性能云存储系统后,整个系统能够保持较优的存储性能和计算性能。

从图 4~6 可以看出,黑色折线表明 HPSCS2A 的性能是尽可能接近系统中性能最优秀的 DFS 系统的性能,说明在基于 HPSCS2A 架构的云存储系统中,系统的整体性能可以在 HPSCS2A 的协调控制下智能地选择最优的 DFS,不能超过系统中最优 DFS 的性能是因为 HPSCS2A 不是一种 DFS 的实现而是一种基于多 DFS 的高层体系架构。

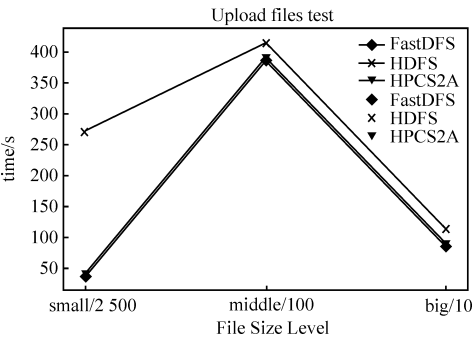


图 4 文件上传时间对比图

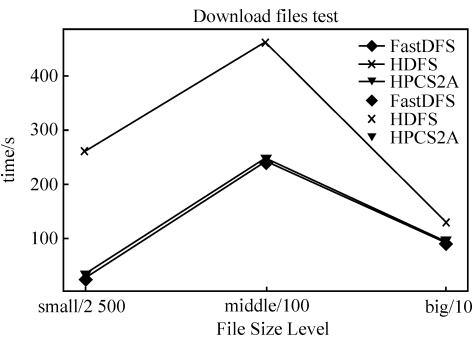


图 5 文件下载时间对比图

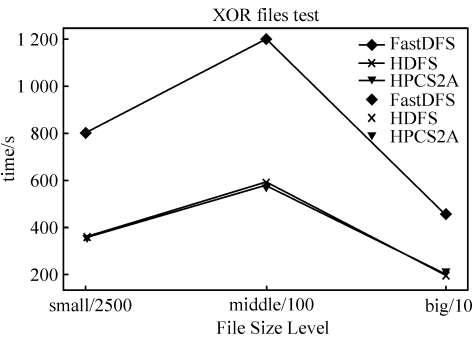


图 6 计算能力时间对比图

4 结束语

随着数据量的激增,云存储技术应运而生^[11-14]。为了提高系统性能,本文提出一种层次式

云存储架构 HPSCS2A。该架构是由多种分布式文件系统共同组成的大系统,并结合这些系统的优点展示出综合性能。实验结果显示,该架构通过对多种 DFS 进行协调和优选,使得存储数据的访问和计算都能达到高效率。

参考文献:

[1] 张继平. 云存储解析[M]. 北京:人民邮电出版社,2013.

[2] FAN KAI, ZHAO LI-BIN, LI HUI, et al. Cloud Storage Architecture with Meta-Data Service Layer in Cloud Computing[C]//The Proceedings of the Third International Conference on Communications, Signal Processing and Systems, Lecture Notes in Electrical Engineering. Nanjing, China, Springer, Cham, 2015 (322):659-667.

[3] 王意洁,孙伟东,周松,等. 云计算环境下的分布存储关键技术[J]. 软件学报, 2012, 23(4):962-986.

[4] 马建红,霍振奇. 基于 HDFS 的创新知识云平台存储架构的研究与设计[J]. 计算机应用与软件, 2016, 33(3):62-66.

[5] 涂序彦. 大系统控制论[M]. 北京:国防工业出版社,2003.

[6] MAI ALFAWAIR. A Cloud Storage Architecture for High Data Availability, Reliability, and Fault-tolerance[C]// Proceedings of the International Conference on Future Networks and Distributed Systems, ICFNDS, 2017. Cambridge, UK, 2017:1-6.

[7] 史策. 云存储与云存储平台搭建技术研究[D]. 大连:大连交通大学, 2017.

[8] KAI FAN, ZHAO LI-BIN, LI HUI, et al. Cloud Storage Architecture with Meta-Data Service Layer in Cloud Computing[C]//The Proceedings of the Third International Conference on Communications Signal Processing, and Systems 2015. Nanjing, China, Springer, 2015(322):659-667.

[9] 王明. 面向小文件的分布式文件存储管理系统的设计与实现[D]. 北京:北京邮电大学, 2018.

[10] CHEN R, DOUGLIS F. Cloud Storage(Guest editors' introduction) [J]. IEEE Internet Computing, 2016, 20(3): 7-9.

[11] 金顺福,王晨飞,陈玲玲,等. Cloud-P2P 云存储结构的模型建立与性能分析[J]. 通信学报, 2015, 36(3): 33-40.

[12] SEN JAYDIP. Cloud Computing - Architecture and Applications[M]. InTech, 2017.

(下转第 32 页)

target=0.033 时会有 0.1 的改善,从而可以有效加强 28 nmPMOS 器件性能.

参考文献:

[1] 张汝京. 纳米集成电路制造工艺 [M]. 2 版. 北京:清华大学出版社, 2017.

[2] KOMODA T, OISHI A, SANUKI T, et al. Mobility improvement for 45nm node by combination of optimized stress and channel orientation design[C]//IEDM Technical Digest. IEEE International Electron Devices Meeting, 2004. New York, IEEE, 2004: 217-220.

[3] 施敏. 半导体器件物理 [M]. 3 版. 西安:西安交通大学出版社,2008.

[4] PARK M S, LEE K T, KANG C Y, et al. The Effect of a Si Capping Layer on RF Characteristics of High- k Metal Gate SiGe Channel pMOSFETs[J]. IEEE Electron Device Letters, 2010, 31(10): 1104-1106.

[5] MENG X, LI Q, HUANG J, et al. Method of improving dislocation for SiGe EPI process[C]//2015 China Semiconductor Technology International Conference. Shanghai, IEEE, 2015: 1-3.

[6] WANG L L, ZHANG J C, JIANG Y L. Optimization of Ni (Pt)/Si-cap/SiGe silicidation for pMOS source/drain contact[J]. IEEE Transactions on Electron Devices, 2017, 64(5): 2067-2071.

[7] OUYANG Q, CHEN X D, MUDANAI S, et al. Two-dimensional bandgap engineering in a novel Si-SiGe pMOSFET with enhanced device performance and scalability[C]//2000 International Conference on Simulation Semiconductor Processes and Devices (Cat. No. 00TH8502). Seattle, WA, USA ,IEEE, 2000: 151-154. .

[8] BAUDOT S, ANDRIEU F, WEBER O, et al. Fully depleted strained silicon-on-insulator p-MOSFETs with recessed and embedded silicon-germanium source/drain [J]. IEEE Electron Device Letters, 2010, 31(10): 1074-1076.

作者简介:

李 帅 男,(1992-),硕士. 研究方向为微电子器件与工艺. E-mail:lee251113@126. com.

蔡小五 男,(1979-),博士,正高级工程师. 研究方向为智能功率集成电路设计、智能高边功率开关、体硅和 SOI 高压半桥驱动电路设计、低边驱动电路设计、体硅 BCD 和 SOIBCD 工艺研究、功率器件 VDMOS 设计、集成电路 ESD 保护设计.

隋振超 男,(1982-),硕士,正高级工程师. 研究方向为工艺整合与微电子器件.

(上接第 27 页)

[13] SU CUI-CUI, WANG YONG-ZHI, SHEN YU-LONG, et al. Improving Database Storage Usability with the Cloud-Based Architecture[C]//IEEE Second International Conference on Data Science in Cyberspace IEEE Computer Society. Shenzhen, China, 2017.

[14] 朱晓彦,顾旭峰. 云存储技术与应用[M]. 北京:高等

教育出版社. 2018.

作者简介:

李海荣 女,(1976-),硕士,副教授. 研究方向为云存储. E-mail:xiaofang611@126. com.

方中纯 男,(1975-),博士,副教授. 研究方向为人工智能.