基于 Ceph 对象存储集群的负载均衡设计与实现①

杨 飞[1], 朱志祥[2], 梁小江[2]

[1](西安邮电大学, 西安 710061)
[2](陕西省信息化工程研究院, 西安 710061)

摘 要: 海量并行数据访问 ceph 对象存储集群时, 会出现访问数据错误率增加和访问速率降低的问题. 首先设计和部署 ceph 对象存储集群, 根据用户请求数设计对象网关节点, 实现用户数据的交互功能, 然后在服务节点安装和部署 haproxy, 实现服务节点的负载均衡功能, 能够降低服务节点的压力, 最后设计和实现四种基于 ceph 对象网关节点的存储集群. 通过大量对比测试, 本设计方案的数据访问错误率降低 0.96%, 用户数据的访问速率提升 74.04%.

关键词: Ceph; 对象网关; 负载均衡; 错误率; 访问速率

Design and Implementation of Load Balancing Based on Ceph Object Storage Cluster

YANG Fei[1], ZHU Zhi-Xiang[2], LIANG Xiao-Jiang[2]

[1](Xi'an University of Posts and Telecommunications, Xi'an 710061, China)
[2](Shaanxi Information Engineering Research Institute, Xi'an 710061, China)

Abstract: When mass parallel data access the ceph object storage cluster, the problems of increasing error rate and low access rate are existed. At the first, the authors design and deploy the ceph object storage cluster, and then design object gateway nodes to achieve interactive features of user data according to user requests. Second, I install and deploy haproxy on four service nodes, which can achieve the function of load balancing and reduce the pressure of four service nodes. Finally, the authors design and implementation of four different kinds of storage clusters based on ceph object gateway nodes. Through a lot of comparison tests, this design can not only reduce the error rate of 0.96 percent, but also enhance access rate of 74.04 percent.

Key words: ceph; object gateway; load balance; error rate; access rate

1 引言

随着云计算与大数据的不断发展, 对于网络数据的存储和处理能力提出新的要求, 当前的数据存储系统已无法满足海量增长的网络数据[1]. Ceph 分布式文件系统集群能够根据海量数据的增长而进行集群扩展, 有效降低服务器并行访问的压力, 提高数据访问的速率, 降低数据访问的错误率[2].

本文以 ceph 分布式文件系统为研究对象, 首先设计部署 ceph 存储集群[3], 结合 mysql 与 keystone 实现 ceph 对象网关节点的统一认证. 然后设计 ceph 对象存储集群的对外网关功能, 实现服务器与 ceph 对象存储集群的数据交互, 能够根据访问服务器用户量的增长, 而扩展 ceph 对象存储集群的对外网关节点, 实现 ceph 对象存储集群的多区域数据管理, 通过负载均衡降低服务器并行数据的压力, 从而提升服务器并行数据访问速率, 同时能够降低用户访问数据的错误率[4].

2 整体设计框架

2.1 设计部署

本设计方案中包括 11 个节点服务器: east、west、south、north 为 ceph 对象存储集群的对外网关节点, 通过 haproxy 负载均衡节点对 east、west、south、north 四个对象网关节点进行负载均衡, node1、node2、node3、node4 是 ceph 对象存储集群的存储和元数据节

① 收稿时间:2015-08-14;收到修改稿时间:2015-09-17

点, 而 deploy 节点为 ceph 集群部署节点, DNS 节点为 ceph 对象存储集群的统一认证中心, haproxy 节点为负载均衡节点, 采用轮询的方式减轻用户访问对象网关节点的压力. 表 1 为设计部署说明表.

表 1 设计部署说明表

节点	角色	安装项目
DNS	token 认证	mysql+keystone
east	gateway	mysql keystone mon
west	gateway	mysql keystone mon
south	gateway	mysql keystone mon
north	gateway	mysql keystone mon
node1	存储节点	ceph osd mds
node2	存储节点	ceph osd mds
node3	存储节点	ceph osd mds
node4	存储节点	ceph osd mds
deploy	管理节点	ceph-deloy
Haproxy	服务节点	haproxy

2.2 设计框架

并行用户数据通过 ceph 对象存储集群的 haproxy 节点将海量并行数据轮询分发到 east、west、south、north 四个 ceph 对象网关节点, 实现并行数据的负载均衡[5]. 在四个对象网关节点中设计和部署 mysql 和 keystone 实现 ceph 对象存储集群的统一认证[6]. 用户可以对 ceph 对象存储集群进行数据操作和管理. 图 1 为整体设计框架.

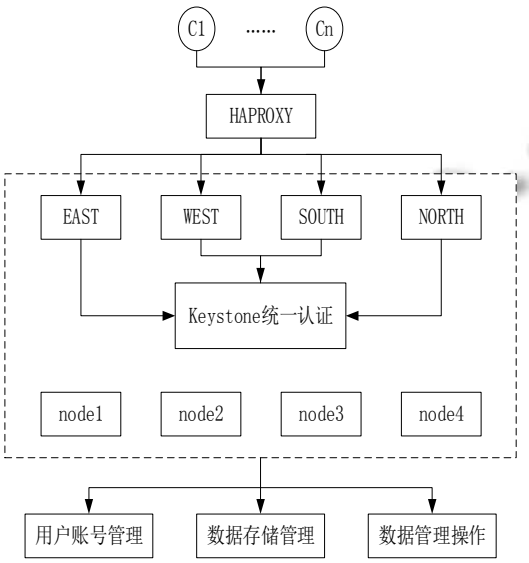

图 1 整体设计框架

在 east、west、south、north 对象网关节点通过统一认证后, 在 haproxy 节点通过负载均衡降低服务器并行数据的压力, 从而有效地提升服务器并行数据访问速率, 同时能够降低用户访问数据的错误率.

3 统一认证和负载均衡

3.1 统一认证

首先在 DNS、east、west、south、north 节点上面安装配置 mysql. 修改 mysql 配置文件, 允许客户端的访问 mysql 数据库, 然后在 mysql 中创建 keystone 数据库, 安装配置 keystone, 对 keystone 进行初始化, 并重启 mysql 和 keystone 的服务.

修改每个对象网关节点的 ceph.conf 配置文件, 添加 keystorgw_keystone_url = 10.10.10.61:5000 . 这样对象网关节点可以使用 DNS 节点产生的统一 token, 实现了所有对象网关节点的统一认证.

首先在 haproxy 节点启动负载均衡的服务, 然后在 DNS 节点产生所有对象网关节点的统一认证的 token.

在 DNS 节点输入命令:DNScurl -d '{"auth": {"tenantName": "admin","passwordCredentials":{"uS ername": "admin","password":"hastexo"}}}' –H "Con tent -type:application/json" http://10.10.10.61:5000/v2 .0/tokens | Python –mjson.tool

在 east、west、south、north 对象网关节点使用命令进行统一认证:

curl -v -H 'X-Auth-Token:token' http://10.10.10.7 1:8888/swift/v1

3.2 负载均衡

在 haproxy 节点设置和部署 ceph 对象存储集群的负载均衡服务. 配置和修改 haproxy.cfg 文件:

listen web_proxy 0.0.0.0: 8888

balance roundrobin

server east 10.10.10.62:8080 cookie 1 weight 5 check inter 2000 rise 2 fall 3

server west 10.10.10.63:8080 cookie 1 weight 5 check inter 2000 rise 2 fall 3

server south 10.10.10.64:8080 cookie 1 weight 5 check inter 2000 rise 2 fall 3

server north 10.10.10.65:8080 cookie 1 weight 5 check inter 2000 rise 2 fall 3

listen stats :8899

负载均衡节点的 IP:10.10.10.71, 负载均衡服务端

口为 8888. 监控界面的端口为 8899, 所有对象网关节点的操作都会通过 10.10.10.71:8888 进行均衡均衡. 图 2 为负载均衡登陆界面[7].

图 2　负载均衡登录界面

在 ceph 对象存储集群中, 设计和部署四种不同的 ceph 对象存储架构图. 在对象网关节点设计四种不同的负载均衡模式, 分别为表 2 单个对象网关、表 3 两个对象网关、表 4 三个对象网关、表 5 四个对象网关.

表 2　单个对象网关

HAPROXY	Queue			Session rate			Sessions					Bytes	
	Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	LbTot	In	Out
East	0	0	-	46	79		76	83	-	73	87	867	938

表 3　两个对象网关

HAPROXY	Queue			Session rate			Sessions					Bytes	
	Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	LbTot	In	Out
East	0	0	-	56	69		76	823	-	783	817	8267	9368
West	0	0	-	79	94		98	1046	-	692	745	7009	9270

表 4　三个对象网关

HAPROXY	Queue			Session rate			Sessions					Bytes	
	Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	LbTot	In	Out
East	0	0	-	24	344		0	66	-	736	715	8362	9467
West	0	0	-	24	343		0	68	-	712	702	8267	9266
South				25	343		106	107		772	730	7634	8713

表 5　四个对象网关

HAPROXY	Queue			Session rate			Sessions					Bytes	
	Cur	Max	Limit	Cur	Max	Limit	Cur	Max	Limit	Total	LbTot	In	Out
East	0	0	-	12	124		10	180	-	272	267	1574	3428
West	0	0	-	11	102		2	116	-	208	287	1493	3705
South	0	0		11	131		2	127		227	195	1354	2920
North	0	0		11	64		3	149		216	177	1304	2856

在 haproxy 负载均衡节点上配置 haproxy.cfg, 然后启动 haproxy 的服务, 在负载均衡监控界面可以对不同设置的对象网关节点进行监控.

4　应用与测试

4.1　测试数据

本方案使用 webbench 并行测试软件对整个 ceph 对象存储集群进行压力测试[8], 测试时间为 120s, 通过大量对比测试, 能够得出海量并行数据访问服务器的失败率和数据速率. 表 6 为单个对象网关测试数据、表 7 为两个对象网关测试数据、表 8 为三个对象网关测试数据、表 9 为四个对象网关测试数据.

表 6　单个对象网关测试数据

测试	并行数	成功	失败	速率
单个对象网关	128	43169	0	0.43
	256	44136	0	0.44
	512	46577	0	0.46
	1024	54894	0	0.48
	2048	62690	164	0.51
	4096	64441	228	0.53

表 7　两个对象网关测试数据

测试	并行数	成功	失败	速率
两个对象网关	128	42823	0	0.46
	256	43925	0	0.49
	512	44874	0	0.52
	1024	58390	18	0.53
	2048	56570	63	0.49
	4096	69269	401	0.60

表 8　三个对象网关测试数据

测试	并行数	成功	失败	速率
三个对象网关	128	41950	0	0.42
	256	52871	0	0.50
	512	53830	0	0.51
	1024	58863	129	0.57
	2048	67386	307	0.66
	4096	69204	638	0.72

表 9　四个对象网关测试数据

测试	并行数	成功	失败	速率
四个对象网关	128	68400	0	0.67
	256	64316	0	0.63
	512	61257	0	0.76
	1024	68439	0	0.77
	2048	80741	0	0.98
	4096	94747	0	1.15

4.2 分析结果

对测试的数据结果进行分析和计算, 从而清晰的得出负载均衡对整个ceph对象存储集群的性能优化情况, 包括并行访问数据的速率和访问数据的错误率. 图 3 为数据访问速率测试对比图, 图 4 为数据错误率测试对比图.

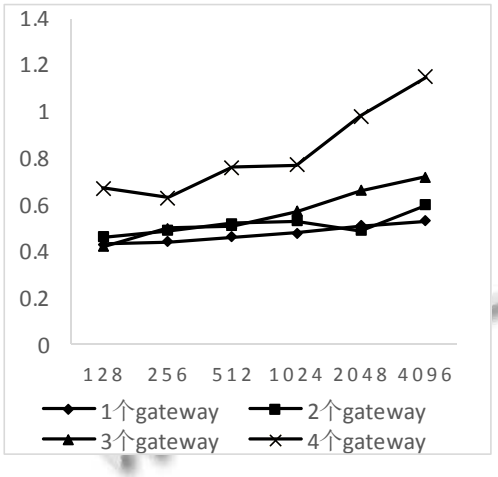

图 3 数据访问速率测试对比图

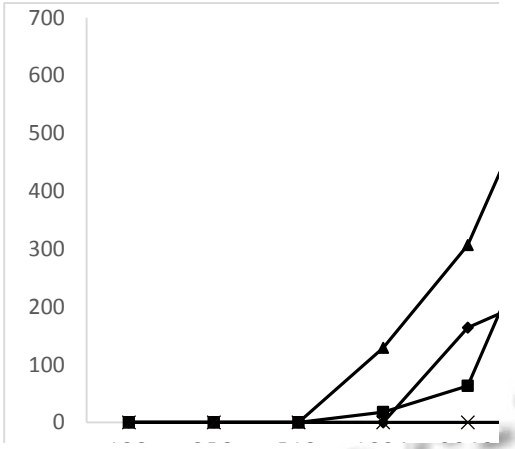

图 4 数据错误率测试对比图

5 总结

设计和部署 ceph 存储集群, 本设计方案能够根据用户的访问请求数, 动态地设计和实现基于 ceph 对象存储集群的对象网关节点, 从而不断适应网络数据的快速增长. 同时在不同的对象网关节点采用统一的认证中心, 确保用户数据的安全性和封闭性.

通过负载均衡和统一认证测试, 本设计能够有效的降低海量并行数据访问服务器的压力, 同时保证用户数据的安全性. 并行访问速率提升 74.04%, 访问数据的错误率降低 0.96%.

参考文献

1 吴广君,王树鹏,陈明,等.海量结构化数据存储检索系统.计算机研究与发展,2012,49.

2 冯幼乐,朱六璋.CEPH 动态元数据管理方法分析与改进.电子技术,2010,47(9).

3 李翔,李青山,魏彬.Ceph 分布式文件系统的究及性能测试. 2014(5):1–15

4 Weil SA, Brandt SA, Miller EL, et al. Ceph: A scalable, high-performance distributed file system. Proc. of the 7th Symposium on Operating Systems Design and Implementation (OSDI). 2006. 307–320.

5 邹仁明,彭隽,李军.OpenStack 开源云平台高可用架构的设计与实现.中国计算机用户协会网络应用分会 2014 年第十八届网络新技术与应用年会 2014.

6 Tang B, Sandhu R. Extending openStack access control with domain trust. Network and System Security. Springer International Publishing, 2014: 54–69.

7 Liu K. To achieve load-balance of elective system with haproxy. Computer Knowledge & Technology, 2011.

8 Yan CR, Shen JY, Peng QK, et al. A throughput-driven scheduling algorithm of differentiated service for web cluster. Wuhan University Journal of Natural Sciences, 2006, 11(1): 88–92.