

第2题（案例题）：

阅读以下关于Web应用系统集成的问题，在答题纸上回答问题1至问题3。

国内的一个某公司为国内某省的某个电信运营商开发了一套基于WEB的全省集中的业务稽核系统，由于该平台的业务数据量在不断增加，导致系统的处理速度越来越明显不能支持新业务扩展的需要。面对这一情况，公司与客户召开了讨论会，寻求该平台的解决方案。公司认为客户应该为各种服务器进行扩容，客户在分析各种服务置之后，表示每台服务器都已经不能再扩容了。最终在讨论会上达成共识，在硬件上和软件上进行双重改进来完成系统的设计方案。也就是客户利用现有遗留的服务器为务器建立集群服务，改进方案如下图5-1所示。另外公司也将从软件技术进行升级，通过缓存技术来实现大批量静态数据的查询。

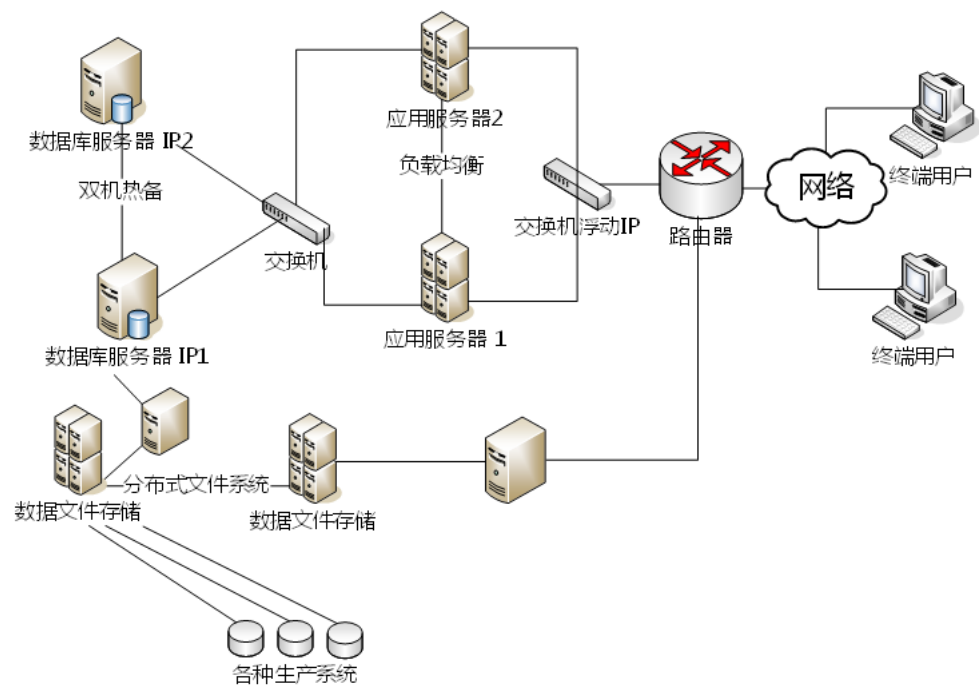


图 5-1 系统改进方案图

【问题1】（10分）
公司通过对该平台的数据经过分析后，发现静态的用户表是一个按月存储的宽表，数据量特别大，而且各种业务的处理都需要大量访问该表，所以，为了提高系统的性能需要使用分布式缓存技术，通过对分布式缓存的Redis技术和Memcached技术进行比较后，最终选择了Redis技术来实现这种分布式缓存服务器。从A-J中分别选出答案填入表5-1中的（1）-（10）处。

表 5-1 分布式缓存技术的比较

比较类别	Redis	Memcached
支持的数据类型	(1) (2) (3) (4)、sorted set、string	(4)
线程	(5)	(6)
数据存储	(7)	(8)
数据备份与恢复	(9)	(10)

- A) 数据丢失后不可恢复
B) 支持单线程

- C) 不支持数据持久化
- D) 支持数据的持久化
- E) Key-value
- F) 哈希
- G) 集合
- H) 支持多线程
- I) 列表
- J) 数据丢失后可恢复, 支持数据的备份

【问题2】(9分)

通过把应用服务器从一台变成了两台, 把用户的请求分散到不同的服务器中, 从而提高负载能力。请给出三种实现应用服务器的软件负载均衡的解决方案, 并简要说明工作机制。

【问题3】(6分)

请用200字以内的文字解释分布式文件系统的概念, 并列2种常见的分布式文件系统解决方案的名称。

选择文件

标记 收藏

【问题1】

(1) F) (2) G) (3) I) (4) E) (5) B) (6) H) (7) D) (8) C) (9) J) (10) A)

其中 (1) (2) (3) 的答案是F)G)I),以上三个答案顺序可调换。

【问题2】

常见的软件负载均衡的实现方案有: http重定向、反向代理服务器、DNS域名解析负载均衡等。

1、http重定向。HTTP重定向就是应用层的请求转发。用户的请求其实已经到了HTTP重定向负载均衡服务器, 服务器根据算法要求用户重定向, 用户收到重定向请求后, 再正的集群。

2、反向代理服务器。在用户的请求到达反向代理服务器时 (已经到达网站机房), 由反向代理服务器根据算法转发到具体的服务器。常用的apache, nginx都可以充当反向器。

3、DNS域名解析负载均衡。DNS域名解析负载均衡就是在用户请求DNS服务器, 获取域名对应的IP地址时, DNS服务器直接给出负载均衡后的服务器IP。

【问题3】

分布式文件系统是指文件系统管理的物理存储资源不一定直接连接在本地节点上, 而是通过计算机网络与节点相连。分布式文件系统的设计基于客户机/服务器模式。一个典可能包括多个供多用户访问的服务器。另外, 对等特性允许一些系统扮演客户机和服务器的双重角色。例如, 用户可以“发表”一个允许其他客户机访问的目录, 一旦被访问, 对客户机来说就像使用本地驱动器一样。常见的解决方案有GFS、HDFS、Lustre、Ceph、GridFS、mogileFS、TFS、FastDFS等。

所属知识点: 案例分析>系统分析

试题难度: 本题共被作答282次 平均得分16.1

参考解析： 本题考查Web系统架构设计的相关知识及应用。

此类题目要求学生认真阅读题目，根据实际系统的需求描述，进行Web系统架构的设计。

【问题1】

本问题考查Web系统设计中分布式缓存的常用技术及实现这些技术的常用软件。

Memcached 是一个高性能的分布式内存对象缓存系统，用于动态Web应用以减轻数据库负载。它通过在内存中缓存数据和对象来减少读取数据库的次数，从而提高动态、数据库驱动网站的速度。Memcached基于一个存储键/值对的hashmap。其守护进程（daemon）是用C写的，但是客户端可以用任何语言来编写，并通过Memcached协议与守护进程通信。

Redis是一个开源的使用ANSI C语言编写、支持网络、可基于内存亦可持久化的日志型、Key-Value数据库，并提供多种语言的API。

Redis是一个key-value存储系统。和Memcached类似，它支持存储的value类型相对更多，包括string(字符串)、list(链表)、set(集合)、zset(sorted set --有序集合)和hash（哈希类型）。这些数据类型都支持push/pop、add/remove及取交集并集和差集及更丰富的操作，而且这些操作都是原子性的。在此基础上，Redis支持各种不同方式的排序。与Memcached一样，为了保证效率，数据都是缓存在内存中。区别的是Redis会周期性的把更新的数据写入磁盘或者把修改操作写入追加的记录文件，并且在此基础上实现了master-slave(主从)同步。Redis支持主从同步。数据可以从主服务器向任意数量的从服务器上同步，从服务器可以是关联其他从服务器的主服务器。这使得Redis可执行单层树复制。存盘可以有意无意的对数据进行写操作。由于完全实现了发布/订阅机制，使得从数据库在任何地方同步树时，可订阅一个频道并接收主服务器完整的消息发布记录。同步对读取操作的可扩展性和数据冗余很有帮助。

【问题2】

本题考查考生对负载均衡技术的相关概念的了解。

负载均衡建立在现有网络结构之上，它提供了一种廉价有效透明的方法扩展网络设备和服务器的带宽、增加吞吐量、加强网络数据处理能力、提高网络的灵活性和可用性。

负载均衡其意思就是分摊到多个操作单元上进行执行，例如Web服务器、FTP服务器、企业关键应用服务器和其它关键任务服务器等，从而共同完成工作任务。

Web系统设计中的负载均衡可以分为软件负载均衡和硬件负载均衡。

软件负载均衡解决方案是指在一台或多台服务器相应的操作系统上安装一个或多个附加软件来实现负载均衡，如http重定向、DNS域名解析负载均衡、反向代理服务器、LVS、Nginx和HAProxy等，它的优点是基于特定环境，配置简单，使用灵活，成本低廉，可以满足一般的负载均衡需求。软件解决方案缺点也较多，因为每台服务器上安装额外的软件运行会消耗系统不定量的资源，越是功能强大的模块，消耗得越多，所以当连接请求特别大的时候，软件本身会成为服务器工作成败的一个关键；软件可扩展性并不是很好，受到操作系统的限制；由于操作系统本身的Bug，往往会引起安全问题。

硬件负载均衡解决方案是直接服务器和外部网络间安装负载均衡设备，这种设备通常称之为负载均衡器，由于专门的设备完成专门的任务，独立于操作系统，整体性能得到大量提高，加上多样化的负载均衡策略，智能化的流量管理，可达到最佳的负载均衡需求。

负载均衡器有多种多样的形式，除了作为独立意义上的负载均衡器外，有些负载均衡器集成在交换设备中，置于服务器与Internet链接之间，有些则以两块网络适配器将这一功能集成到PC中，一块连接到Internet上，一块连接到后端服务器群的内部网络上。

一般而言，硬件负载均衡在功能、性能上优于软件方式，不过成本昂贵。比较主流的几类产品如F5、Radware等。

【问题3】

本题考查分布式文件系统的相关概念。

分布式文件系统是指文件系统管理的物理存储资源不一定直接连接在本地节点上，而是通过计算机网络与节点相连。分布式文件系统的设计基于客户机/服务器模式。一个典型的网络可能包括多个供多用户访问的服务器。另外，对等特性允许一些系统扮演客户机和服务器的双重角色。例如，用户可以“发表”一个允许其他客户机访问的目录，一旦被访问，这个目录对客户机来说就像使用本地驱动器一样。

计算机通过文件系统管理、存储数据，而信息爆炸时代中人们可以获取的数据成指数倍的增长，单纯通过增加硬盘个数来扩展计算机文件系统的存储容量的方式，在容量大小、容量增长速度、数据备份、数据安全等方面的表现都差强人意。分布式文件系统可以有效解决数据的存储和管理难题：将固定于某个地点的某个文件系统，扩展到任意多个地点/多个文件系统，众多的节点组成一个文件系统网络。每个节点可以分布在不同的地点，通过网络进行节点间的通信和数据传输。人们在使用分布式文件系统时，无需关心数据是存储在哪个节点上、或者是从哪个节点从获取的，只需要像使用本地文件系统一样管理和存储文件系统中的数据。

到目前为止，有很多种分布式文件系统解决方案可供选择，常见的分布式文件系统有，GFS、HDFS、Lustre、Ceph、GridFS、mogileFS、TFS、FastDFS等，他们对使用的环境和适应的文件形式各不相同，加速效果也不相同。在国内mogileFS较为受欢迎，在图片存储上性能突出。而淘宝网开发出了TFS的文件系统，主要针对海量的非结构化数据，它建立在普通的Linux机器集群上，可为外部提供高可靠和高并发的存储访问。

1、GFS:是由google开发并设计的一个面向大规模数据处理的一个分布式文件系统.为了满足Google迅速增长的数据处理需求，Google设计并实现了Google文件系统。它是有几百甚至几千台普通的廉价设备组装的存储机器。GFS采用主/从模式，一个GFS包括一个master服务器和多个chunk服务器。当然这里的一个master是指逻辑上的一个，物理上可以有多个（就是可能有两台，一台用于以防万一，一台用于正常的数据管理）。

2、HDFS以流式数据访问模式来存储超大文件，它是谷歌的GFS提出之后出现的另外一种文件系统，它有一定高度的容错性，而且提供了高吞吐量的数据访问，非常适合大规模数据集上的应用。

3、Lustre是HP、Intel、Cluster File System公司联合美国能源部开发的Linux集群并行文件系统，是第一个基于对象存储设备的，开源的并行文件系统。它由客户端，两个MDS，OSD 设备池通过高速的以太网或 QWS Net 所构成。可以支持1000 个客户端节点的 I/O 请求，两个 MDS 采用共享存储设备的 Active - Standby方式的容错机制，存储设备跟普通的，基于块的 IDE 存储设备不同，是基于对象的智能存储设备。

4、GridFS是mongoDB自带的文件系统，它以二进制的形式存储文件 大型文件系统的绝大多数特性GridFS全可以完成。GridFS会直接利用已建立的复制、分片机制；GridFS可以避免传统文件系统的某些弊端，例如同一目录下文件数量过多；GridFS不会产生磁盘碎片；

笔记

提问

上一题

查看答案与解析

下一题

