哈尔滨工业大学

**计算机科学与技术学院/国家示范性软件学院**

**2020年春季学期**

**《软件架构与中间件》课程**

**实验报告**

**Lab 2：计算层软件架构实验**

|  |  |  |
| --- | --- | --- |
| **姓名** | **学号** | **联系方式** |
| 黄项宇轩 | 1173710106 | 1171845717@qq.com |
| 任永鹏 | 1173710107 | 805402160@qq.com/18800420869 |

目 录

[1 实验概述 1](#_Toc38980088)

[1.1 实验目的 1](#_Toc38980089)

[1.2 实验要求 1](#_Toc38980090)

[2 实验内容与过程 1](#_Toc38980091)

[2.1 nginx集群负载均衡实验 1](#_Toc38980092)

[2.2 hadoop分布式计算实验 13](#_Toc38980093)

[2.3 activeMQ异步消息推送实验 23](#_Toc38980094)

[3 结对开发过程记录 23](#_Toc38980095)

[4 实验总结 24](#_Toc38980096)

[5 教师评语 25](#_Toc38980097)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **学号：** | **1173710106** | **姓名：** | **黄项宇轩** | |
| **学号：** | **1173710107** | **姓名** | **任永鹏** | |
| 实验概述 | | | |  |
| 实验目的 1）学习构建集群的方法  2）学会使用典型的负载均衡框架  3）学会使用Hadoop 进行map-reduce编程  4）能够灵活应用计算层中间件到实际系统 实验要求 1）2人结对成组  2）实验2.1和2.2必做，实验2.3选做  3）结合《软件过程与工具》课程中进销存系统(或其他实际软件系统)进行计算层架构重构，支持海量用户的在线高并发请求场景  4）应给出关键过程的细节 | | | | |
| 实验内容与过程 | | | |  |
| nginx集群负载均衡实验  1. 请给出搭建单虚拟机nginx的关键过程，并修改nginx主机的首页内容从英文变为中文。   安装Nginx  sudo apt-get update  sudo apt-get install nginx    查看版本，安装成功  Nginx -v    修改主机host文件，配置域名        2）请详析配置文件中涉及的upstream、location和server块。    如上图，master结点，server模块，listen设置监听接口，server\_name设置本机ip地址，  Location模块设置配置proxy\_pass代理转发时的路径node,下面为存放代理的请求头相关参数，upstream模块，ip\_hash设置负载均衡算法，同时设置两个后端服务器的ip地址。    上图为slaver结点，同理，server模块listen设置监听接口，server\_name设置本节点ip地址，location模块，root方式设置资源路径。  3）请描述所搭建的虚拟集群环境，验证虚拟集群可以互相访问。    分别在两个slaver node文件夹中放入不同的index.html文件，然后通过访问master结点ip，通过负载均衡可以进行访问其他节点的页面  4）配置集群的负载均衡系统，给出关键过程，验证负载均衡系统是否工作。  如图，默认为weight=2，即轮询算法，在浏览器中输入192.168.174.129，根据轮询算法，负载均衡系统将会均匀分配到192.168.174.130和192.168.174.131两个后端服务器当中      5）配置并验证使用不同集群负载均衡算法（至少三种），给出过程、观测的情况，给出自己对负载均衡算法的理解和认识。  算法1：轮询  轮询是默认的方式，每个请求按时间顺序逐一分配到不同的后端服务器上。如果后台服务器上某一台宕机了，它可以自动剔除。  特点为权重相同，结果如下，其中分配访问两个服务器的次数基本相同。      算法2：加权轮询  指定轮询的几率，wight和访问比率成正比，用于后台服务器性能不均匀的情况。  这里我设置权重差别比较大，效果也很明显，访问两个服务器的次数有着明显差异。Wight越高，访问比率越高。    算法3：ip\_hash  根据每个请求的ip的hash结果分配，因此每个固定ip能访问到同一个后端服务器，可以解决session问题。设置之后，基本访问的是固定一个服务器。    算法4：fair  按照后端服务器的相应时间来分配请求，时间短的优先分配。根据该算法进行配置，效果为基本固定访问同一个后端服务器。    6） 应用负载均衡技术改造遗留的“进销存”系统，赋予支持海量用户的在线高并发请求的能力，请给出设计细节并分析负载均衡前后的区别。  首先将进销存系统打包成jar包，然后通过共享文件夹，传入到master结点中，之后使用java -jar app.jar方法，进行测试localhost:8081页面是否可以打开    然后修改master结点的配置文件，slaver结点端口改为8081    在其他两个节点启动jar包  然后访问master 的ip，进行负载均衡    接下来进行压力测试  压力测试使用apache bench进行压力测试  命令行为ab -c 10 -n 100 + 测试url  意思为模拟10个用户共发送1000个请求  详细测试数据如下：  服务器软件:nginx / 1.10.3  服务器主机名:192.168.174.129  服务器端口:80  文件路径:/ # /登录? = % 2 findex重定向  文档长度:800字节  并发级别:10  测试时间:1.325秒  完成要求:100  失败的请求:0  Non-2xx反应:100  传输总量:98600字节  传输的HTML: 80000字节  每秒请求数:75.44[#/秒](平均)  每次请求所需时间:132.550 [ms](平均值)  每个请求的时间:13.255 [ms](所有并发请求的平均时间)  传输速率:72.64 [Kbytes/sec]接收  连接次数(Miss)  最小均值[+/-sd]中值最大值  连接:0 0 0.2 0 1  处理:1 111 165.0 23 821  等待:1 111 165.0 23 821  总数:1 111 165.1 23 821  在一定时间内处理的申请的百分比(毫秒)  50% 23  66% 114  75% 172  80% 255  90% 325  95% 350  98% 760  99% 821  100% 821(最长请求)  下面是截图    为比较负载均衡的效果，测试不进行负载均衡的情况下的数据  控制变量同样为10个用户共发送1000个请求  测试数据如下：  服务器软件:  服务器主机名:localhost  服务器端口:8081  文件路径:/ # /  文档长度:800字节  并发级别:10  测试时间:2.528秒  完成要求:100  失败的请求:0  Non-2xx反应:100  传输总量:95500字节  传输的HTML: 80000字节  每秒请求数:39.55[#/秒](平均)  每次请求所需时间:252.833 [ms](平均值)  每个请求的时间:25.283 [ms](所有并发请求的平均时间)  传输速率:36.89 [Kbytes/sec]接收  连接次数(Miss)  最小均值[+/-sd]中值最大值  连接:0 1 4.9 0 35  处理:0 247 603.4 15 2195  等待:0 222 602.2 6 2147  总计:0 248 605.5 15 2195  在一定时间内处理的申请的百分比(毫秒)  50% 15  66% 54  75% 87  80% 145  90% 2014  95% 2015  98% 2051  99% 2195  100% 2195(最长请求)  下面为截图    根据对比测试数据可以发现，使用负载均衡，测试时间，请求时间明显缩短！  7）迁移到真实集群环境重做负载均衡实验(选做)。请给出实验过程中需要注意的问题，比较并分析虚拟集群和真实集群下的异同。 hadoop分布式计算实验 1）配置集群环境，设置主节点主机名为manager，从节点主机名为workerX(X为数字编号)，给出主要过程；同时给出配置SSH免密登录的关键步骤。  修改hostname和hosts      生成rsa公钥私钥    将id\_rsa.pub公钥复制到authorized\_keys中，将授权文件拷贝到slave1和slave2    2）请配置hadoop集群计算环境，要求数据块副本数为3。给出主要过程，并验证正确性。    格式化    出现master: mkdir: 无法创建目录"/sam/hadoop-3.1.2/logs": 权限不够  使用 sudo chown -R 你自己电脑的用户名 "/sam/hadoop-3.1.2/logs 命令行    集群搭建完毕  3）请使用hadoop集群环境计算英文版[Stray Birds](https://en.wikisource.org/wiki/Stray_Birds)正文中每个单词出现的次数。  上传文件  hadoop fs -put StrayBirds.txt hdfs://192.168.174.129:9000/Birds    hadoop jar ./share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar wordcount hdfs://192.168.174.129:9000/Birds hdfs://192.168.174.129:9000/BirdResult    登录9870端口    下载结果    4）请使用hadoop集群环境计算遗留的“进销存”系统中海量用户的log日志中的访问统计。请模拟输入数据，从1万个用户共100万次操作记录中计算每个用户的访问次数。  用户的log日志  结果，第一列为用户id，第二列为访问次数 activeMQ异步消息推送实验 1）安装并配置ActiveMQ 5.15.9，验证正确性。  2）请先创建名为asm和se的主题，然后面向这些主题，给出发送端和接收端的交互流程，最后编写流程对应的代码。  3）请完成如下要求：如果发送的消息是object，该如何处理；如何保证消息处理成功之后，会发送消息确认。  4）请比较实验1自主开发的消息中间件与ActiveMQ的异同，重点给出如何改进自己程序的不足。 | | | | |
| 结对开发过程记录 | | | |  |
| **（1）角色切换与任务分工**  表1-1结对开发角色与任务分工   | 日期 | 时间(HH:MM - HH:MM) | 驾驶员角色 | 领航员角色 | 本段时间的任务 | | --- | --- | --- | --- | --- | | 2020/4/15 | 15:00-18:00 | 黄项宇轩 | 任永鹏 | Nginx集群搭建 | | 2020/4/19 | 15:00-18:00 | 任永鹏 | 黄项宇轩 | Nginx压力测试 | | 2020/4/22 | 15:00-18:00 | 黄项宇轩 | 任永鹏 | Hadoop集群搭建 | | 2020/4/25 | 15:00-18:00 | 任永鹏 | 黄项宇轩 | Hadoop应用与遗留改造 |   **（2）工作日志**  由领航员负责记录，记录结对开发期间的遇到的问题、两人如何通过交流合作解决每个问题的。  表1-2 结对开发工作日志   | 日期/时间 | 问题描述 | 最终解决方法 | 交流过程 | | --- | --- | --- | --- | | 2020/4/15 | Nginx配置无效 | 将默认配置路径注释 | 对比配置无误是别的问题 | | 2020/4/15 | 虚拟机崩溃 | 修改虚拟机占用内存 |  | | 2020/4/22 | Ssh连接文件权限受限 | sudo chown -R sam /usr/local/hadoop | chmod修改无效 |   **（3）结对开发工作现场照片、或视频及文件沟通截图** | | | | |
| 实验总结 | | | |  |
| Nginx同Apache一样都是一种 Web 服务器。基于 REST 架构风格，以统一资源描述符URI 或者统一资源定位符URL 作为沟通依据，通过 HTTP 协议提供各种网络服务。然而，这些服务器在设计之初受到当时环境的局限，例如当时的用户规模，网络带宽，产品特点等局限并且各自的定位和发展都不尽相同。这也使得各个 Web 服务器有着各自鲜明的特点。Nginx 使用基于事件驱动架构，使得其可以支持数以百万级别的 TCP 连接。高度的模块化和自由软件许可证使得第三方模块层出不穷（这是个开源的时代啊）。Nginx 是一个跨平台服务器，可以运行在 Linux、Windows、FreeBSD、Solaris、AIX、Mac OS 等操作系统上。这些优秀的设计带来的极大的稳定性。Nginx 是一款自由的、开源的、高性能的 HTTP 服务器和反向代理服务器；同时也是一个 IMAP、POP3、SMTP 代理服务器。Nginx 可以作为一个 HTTP 服务器进行网站的发布处理，另外 Nginx 可以作为反向代理进行负载均衡的实现。  Hadoop是由java语言编写的，在分布式服务器集群上存储海量数据并运行分布式分析应用的开源框架，其核心部件是HDFS与MapReduce。HDFS是一个分布式文件系统：引入存放文件元数据信息的服务器Namenode和实际存放数据的服务器Datanode，对数据进行分布式储存和读取。　　MapReduce是一个计算框架：MapReduce的核心思想是把计算任务分配给集群内的服务器里执行。通过对计算任务的拆分再根据任务调度器对任务进行分布式计算。Hadoop的框架最核心的设计就是：HDFS和MapReduce。HDFS为海量的数据提供了存储，则MapReduce为海量的数据提供了计算。把HDFS理解为一个分布式的，有冗余备份的，可以动态扩展的用来存储大规模数据的大硬盘。把MapReduce理解成为一个计算引擎，按照MapReduce的规则编写Map计算/Reduce计算的程序，可以完成计算任务。 | | | | | Nginx 同 Apache 一样都是一种 Web 服务器。基于 REST 架构风格，以统一资源描述符（Uniform Resources Identifier）URI 或者统一资源定位符（Uniform Resources Locator）URL 作为沟通依据，通过 HTTP 协议提供各种网络服务。  然而，这些服务器在设计之初受到当时环境的局限，例如当时的用户规模，网络带宽，产品特点等局限并且各自的定位和发展都不尽相同。这也使得各个 Web 服务器有着各自鲜明的特点。 |
| 教师评语 | | | |  |
|  | | | | |