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**

**2022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** |
| XXX | NNNNNNN | Email/手机号码 |
| XXX | NNNNNNN | Email/手机号码 |

目 录

[1 实验概述 1](#_Toc69932094)

[1.1 实验目的 1](#_Toc69932095)

[1.2 实验要求 1](#_Toc69932096)

[2 实验内容与过程 1](#_Toc69932097)

[2.1 nginx集群负载均衡实验 1](#_Toc69932098)

[2.2 hadoop分布式计算实验 2](#_Toc69932099)

[2.3 activeMQ异步消息推送实验 2](#_Toc69932100)

[3 结对开发过程记录 2](#_Toc69932101)

[4 实验总结 3](#_Toc69932102)

[5 教师评语 4](#_Toc69932103)

[文档全部完成之后，请在上述区域点击右键，选择“更新域”，在打开的对话框中选择“更新整个目录”]

|  |  |
| --- | --- |
| 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。  2）请详析配置文件中涉及的upstream、location和server块。  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。  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。  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。  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。  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master，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。   前期配置  安装JDK  下载步骤省略  解压文件到~/jdk/目录    修改/etc/profile文件，配置JDK信息。然后执行source /etc/profile命令生效。    检测是否成功安装    权限设置，为sam用户添加权限    安装ssh    为当前用户sam赋予拥有权限：sudo chown -R sam .ssh    生成RSA密匙对：ssh-keygen -t rsa    集群内共享密匙： cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys    安装并运行Hadoop  下载及解压步骤省略  配置 hadoop 的环境变量    修改Hadoop安装目录/etc/hadoop/目录下的Hadoop-env.sh文件    core-site.xml中主要内容    hdfs-site.xml中主要内容    mapred-site.xml中主要内容    masters文件，如果没有手动添加一个，配置如下    workers文件    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。  前面步骤均在单机上运行，接下来克隆虚拟机，并进行集群配置  修改 /etc/hostname 文件：master节点的主机设置为 master ，其他两个虚拟机分别设置为slave1、slave2        获取三台虚拟机ip，修改 /etc/hosts文件          确认是否可以免密登录，例如ssh slave1    在master节点上执行hdfs namenode -format    在master节点上执行start-all.sh验证正确性      进入hadoop安装目录中的样例程序目录：  列出当前路径下的文件    新建文本文件words作为输入，内容输入如下，然后使用命令上传到hdfs文件系统中      hadoop jar hadoop-mapreduce-examples-3.1.4.jar wordcount /input /output    用命令：hdfs dfs -cat /output/\* 直接查看结果     1. 请使用Hadoop集群环境计算[英文版《Stray\_Birdss》](https://en.wikisource.org/wiki/Stray_Birds)正文中每个单词出现的次数。   首先，删除前面运行造成的遗留文件，这里省略下载《Stray Birds》的步骤    这里是运行结果     1. 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。   生产模拟数据的程序如下，假设其格式为yyyy-MM-dd HH:mm:ss.SSS [info] userId request.  这里直接为每个用户生产100条记录，共计1000000条数据。    这里为生产的数据    为该log文件编写对应的mapreduce程序，其基本原理与word count类似，具体程序如图，这里不再赘述。        同样，删除hdfs中的遗留文件，上传生成的log文件，并执行打包好的程序    如下为输出的执行结果，因为在生成数据时直接为每个用户生成100条数据，结果中每个用户的访问此时都为100。     ActiveMQ异步消息推送实验  1. 安装并配置ActiveMQ 5.15.9，验证正确性。   下载，解压文件，执行启动命令。    查看localhost:8161，使用用户名admin密码admin登录（默认设置，未修改），以下为登陆后的界面       1. 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。   这里先给出比较通用的向某个主题发送消息的过程，    同上，这里给出通用的从某个主题接收消息的过程。    接下来是为了满足实验要求，做的一些封装。构建一个MyPublisher类和一个MySubscriber类，MyPublisher类提供构造方法，绑定一个topic，并可以使用publishMessage方法向中间件发送消息，MySubscriber类提供构造方法，绑定其要订阅的topic，以及subscibe方法供其继续订阅其他topic，另外提供getMessage方法从中间件获取订阅的所有消息。简要代码如下，不再详细解释，具体见附件中的代码。      接下来是两个测试类，可见对Publish的测试中，为两个主题sam，se分别定义声明两个publisher，然后向中间件发消息。对Subscribe的测试中，定义三个subscriber，第一个订阅sam，第二个订阅se，第三个全部都订阅，并全部调用getMessage从中间件获取消息。（必须先调用Subscribe的测试再调用Publish的测试，或者直接使用程序中的Test类中的测试）      如图是测试结果。    网页上也可以看到相应的变化。     1. 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。   如果发送的消息是一个Java对象，需要将它们序列化，即实现Serializable接口，然后将其发送到ActiveMQ队列或主题中。当消息接收者接收到消息时，它必须还原Java对象并对其进行类型检查和解序列化。在序列化 Java 对象时，可以使用一个通用的格式，比如 JSON 或 XML，以确保不同平台之间的互操作性。  为了将Java对象序列化成字节流，也可以使用Java的对象序列化功能或其他第三方序列化框架，如Google Protocol Buffers、Apache AVRO等。由于Java对象序列化存在安全漏洞，也可能会降低性能，因此建议评估不同序列化方法，并选择适合的序列化方式。  在ActiveMQ中，消息确认机制是通过ACK（acknowledge）信号实现的。当一个消费者从一个队列或主题中取出一个消息并成功处理后，该消息就可以被确认了。消费者可以发送ACK信号来确认这个消息被成功处理。  在ActiveMQ中，有两个基本的消息确认模式：   1. 自动确认模式：当消费者从队列或主题中取出消息之后，会自动向ActiveMQ代理发送ACK信号，标志着该消息已经被成功处理。 2. 手动确认模式：消费者必须在处理完消息之后，显式地调用session.acknowledge()方法，以向ActiveMQ代理发送ACK信号。如果没有确认消息，则消息会继续留在队列或主题中，直到下一个消费者取走或过期。   在手动确认模式中，可以在处理消息时选择合适的时机去确认消息，例如只有当消息处理完成且存储到数据库中时再确认消息。这样可以确保在发生异常情况时，消息不会重复消费，从而提高系统的可靠性和稳定性。  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。  相同点：都进行了多模块解耦，  不同点：ActiveMQ 是支持持久化的，可以永久保存消息。消息是保存在内存中的。当内存空间不足，或者ActiveMQ 服务关闭的时候，消息会被持久化到磁盘上。 消息被消费的时候，再加载到内存空间中。  另外，ActiveMQ 支持流量削峰。可以控制活动的人数，可以缓解短时间内高流量压垮应用；  改进如下：在实验一自主开发的中间件的应用前端加入消息队列。用户的请求，服务器接收后，首先写入消息队列。假如消息队列长度超过最大数量，则直接抛弃用户请求或跳转到错误页面；根据消息队列中的请求信息，再做后续处理。 | |
| 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 | | --- | --- | --- | --- | --- | |  |  |  |  |  | |  |  |  |  |  | |  |  |  |  |  |   【注意】该表格可自行增加更多的行。  **（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 | | --- | --- | --- | --- | |  |  |  |  | |  |  |  |  | |  |  |  |  |   【注意】该表格可自行增加更多的行  **（3）结对开发工作现场照片、或视频及文件沟通截图**  请其他同学帮助拍摄结对开发现场照片至少2张。  13071045_wLhN 1  结对开发现场照片1 结对开发现场照片2 | |
| 实验总结 |  |
| 【结合前期课程项目系统或其他实际软件系统，解决其中的消息订阅/分发问题等】 | |
| 教师评语 |  |
|  | |