**北 京 邮 电 大 学**

**本科毕业设计（论文）任务书**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 学院 | 计算机学院（国家示范性软件学院） | | 专业 | 智能科学与技术 | | 班级 | 2019211315 |
| 学生姓名 | 张梓靖 | | 学号 | 2019211379 | | 班内序号 | 27 |
| 指导教师姓名 | 王纯 | | 所在单位 | 计算机学院（国家示范性软件学院） | | 职称 | 高级工程师 |
| 设计(论文)题目 | （中文）一种基于工作量的Serverless计算自动伸缩算法的设计与实现 | | | | | | |
| （英文）Design and Implementation of Workload-based Auto-scaling Algorithm for Serverless Computing | | | | | | |
| 题目分类 | 工程实践类□ 研究设计类☑ 理论分析类□ | | | | | | |
| 题目来源 | 题目是否来源于科研项目 是□ 否☑ | | | | | | |
| 科研项目名称： | | | | | | |
| 科研项目负责人： | | | | | | |
| 主要任务及目标：   * 1、学习Kubernetes容器编排系统的知识，并在虚拟机中搭建一个Kubernetes集群，熟悉Kubernetes的基本操作，并基于该集群环境进行后续实验。重点学习和使用Kubernetes的工作负载资源Deployment和HPA控制器，并理解HPA控制器的基于阈值的自动水平伸缩工作负载实例的算法。 * 2、学习Serverless计算的知识，并在搭建的Kubernetes集群中尝试使用Serverless框架OpenFaaS。 * 3、根据Serverless工作负载流量的突发性，设计一个Serverless服务，其流量按照一定规律随时间而变化，并且可通过OpenFaaS框架部署到搭建好的Kubernetes集群。以设计的Serverless服务为示例应用，设计实现用来预测Serverless工作负载请求流量的时间序列预测算法。 * 4、以设计的Serverless服务为示例应用，设计实现根据预测流量值的自动水平伸缩工作负载实例的算法。 * 5、将算法部署到Kubernetes集群，与OpenFaaS框架进行交互。 | | | | | | | |
| 主要内容：  1、学习和使用Kubernetes容器编排系统。  2、学习并掌握Serverless计算的开源框架OpenFaaS。  3、调研并使用主流时间序列预测方法，根据历史数据来预测Serverless工作负载的未来请求流量情况，以进行工作负载实例数量的自动水平伸缩。  4、部署到Kubernetes集群进行验证。 | | | | | | | |
| 主要参考文献：  [1]Kubernetes Authoritative guide version 4, author: Zheng Gong, Zhihui Wu, Xiulong Cui, Jianyong Yan.  [2]Docker technology introduction, author: Baohua Yang.  [3]Kubernetes docs: https://kubernetes.io/docs/home/  [4]OpenFaaS docs: https://docs.openfaas.com/  [5]Eric Jonas, Johann Schleier-Smith, Vikram Sreekanti and et al., “Cloud programming simplified: A berkeley view on serverless computing,“ arXiv preprint arXiv:1902.03383, 2019.  [6]Laszlo Toka, Gergely Dobreff, Balazs Fodor and Balazs Sonkoly, “Adaptive AI-based auto-scaling for Kubernetes,“ in 2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID). IEEE, 2020, pp. 559-608. | | | | | | | |
| 进度安排：  1、学习和使用Kubernetes容器编排系统，在Kubernetes集群中尝试使用Serverless框架OpenFaaS。2022.11-12  2、基于OpenFaaS设计一个Serverless示例应用，设计实现用来预测Serverless工作负载请求流量的时间序列预测算法，并设计实现根据预测流量值的自动水平伸缩工作负载实例的算法。2023.1-3  3、将设计的算法部署到Kubernetes集群上进行实际测试，获取实验结果，开始撰写论文。2023.4-5  4、修改完善论文，进行答辩。2023.6 | | | | | | | |
| 指导教师签字 | |  | | 日期 | 2022年 12月 9日 | | |