南昌大学校级大学生创新创业训练计划项目

申请书

项目名称 : 基于分布式混合云的大数据可视化系统 □ 创新训练项目 项目类 □ 创业训练项目 型 □ 创业实践项目 学院名称 : 信息工程学院 项目负责人 柳杰 项目组成员 黄嘉诚,邬志文,刘磊,饶隽,熊凯文,黄 智文,罗子涵 导师姓名 · 徐子晨 填表日期 : 2017年5月9日

二O 年 月

大学生创新创业训练计划项目申请书

项目名称			基于分布式混合云的大数据可视化系统						
申请经费			4000		起止时间	2017年 5月至 2018年 5月			
项目类型			□创新	□创新训练项目 □创业训练项目 □创业实践项目					
申请团队		姓名	年级	所在院系、专业	联系电话	E-mail			
	项目负责人及项目组成员	柳杰	15 级	信息工程学院网 络工程	13037239781	442961832@qq. com			
		黄嘉诚	15 级	信息工程学院计 算机科学与技术	18270263352	2277103364@qq.com			
		邬志文	15 级	信息工程学院网 络工程	18279127320	2277487167@qq.com			
		饶隽	16 级	信息工程学院计 算机类	13077950269	799061496@qq.com			
		刘磊	15 级	信息工程学院计 算机科学与技术	13037216993	764813981@qq.com			
		黄智文	15 级	信息工程学院网 络工程	13037206737	2277103364@qq.com			
		熊凯文	16 级	信息工程学院电 子信息类	15979472427	2396361392@qq.com			
		罗子涵	15 级	信息工程学院通 信工程	13635933226	100841836@qq.com			
导师		姓名	徐子晨		职务/职称	教授			
		电话	15797846968		E-mail	xuz@ncu. edu. cn			
		姓名	于海雯		职务/职称	讲师			
		电话	13907007575		E-mail	hoau3qh@dingtalk.com			

一、项目简介(200字以内,包括项目的简要内容以及实施目标、创新点等)

我们的系统的实施目标是面对多维海量数据,构建一个可商业化的分布式数据可视化系统。系统可分为基础分布式计算架构层-采用新兴的 Hadoop 分布式计算架构技术体系,数据计算处理层-采用 Spark,MapReduce,Storm 等新的分布式计算技术,数据可视化表现层-采用丰富的 web 图表形式以不同的方式形象的展示数据信息。我们的系统可以广泛的适用与各类企业信息化管理,如企业人力资源管理,财务管理,销售信息等。2016 年,各行各业的大数据应用都渐渐从空洞的理论落地,被专家们称为"大数据元年",目前数据可视化系统在社会的市场需求前景广阔。

二、申请理由(包括自身具备的知识条件、申请目的等)

团队成员知识条件

知识基础

团队成员都来自信息工程学院,对 C,C++,JAVA,Python 等基本编程语言都有熟练运用,对 Spring,SpringMVC,Mybatis 等基本的 web 开发框架都能熟练运用。在大数据技术架构方面,我们团队成员在导师实验室搭建了分布式的Hadoop 集群,Hbase 分布式面向列存储数据库,分布式缓存数据库 Redis 和Memcached。针对大规模数据处理的 MapReduce 和 Spark 计算框架,针对大规模数据存储管理和查询分析的 Hive 数据仓库,项目开发的知识储备已经充足,可以面对复杂的项目开发任务。

项目经历

团队曾经合作开发个多个WEB项目

2016-2017年:

- 5月: 首次合作开发南昌某高中班级信息管理 WEB 平台, 目前平台一直在正常运维中。
- 7月:为天香熟食批发公司开发集手机 APP, 微信公众号, PC-WEB 端的WEB 进销存系统, 目前项目运行稳定。

2017-2018年

- 3月:合作开发南昌大学本科审核评估教育网,目前项目稳定运行,项目发布后,团队在一起创立 UP-VAMS 工作室。
 - 4月:开发信息工程学院教学数据分析平台。

申请目的

随着企业信息化管理的不断深入,用信息技术提高管理效率、促进业务流程和组织结构的重组和优化,实现资源的优化配置和高效应用,增强了产、供、销协同运作能力,提高企业的市场反应能力、科学决策水平和经济效益,为企业创造了更大价值,但同时也积累的大量的管理数据,如销售数据、生产数据、品质数据、财务数据等,这些海量数据由于没有科学的分析和挖掘工具,使得这些数据价值得不到充分的发挥。

我们的项目主要利用实验室的基础计算设备以及微软 Azure 上的云计算资源搭建来的基于分布式混合云的大数据可视化系统来面对企业的大规模数据分析和数据挖掘需求。本项目的核心目的是帮助企业在两方面提升企业的核心竞争力:

- 1. 通过大规模的数据分析得到可视化的结果,帮助企业建立科学合理的管理模式和管理机制。
- 2. 借助我们在实验室开发的数据分析管理工具,帮助企业收集数据和分析数据,最后实现企业大数据应用。从而持续提升企业的竞争力、决策力和市场预知能力。

三、立项背景(项目研究现状、创新创业前景及意义等,参考关联项目等)项目研究现状

目前针对需求, 我们设计相关的大数据分析系统架构的方案, 并且整个架构的底层开发部署方案已经基本通过测试, 目前团队正在对平台上的项目应用进行开发。

创新创业前景及意义

随着中国改革进入深水区,市场管理不断规范,国家为了扶持中小企业的发展,制定了很多优惠政策,所以中小企业发展遇到空前的机遇,企业的数量和规模都比以前有很大的突破,但同时也变得异常激烈。企业要在竞争中取胜,必须加强企业的内部管理和提升对市场的敏感度和准确把握市场方向,要想实现这两个目标,建立喝血的管理机制和应用数据分析工具是必由之路。

据不完全统计,截至2016年12月,全国小微企业数量为4200万家,中大型企业11500万家。并且以10%的速度增加,按照小微企业每年需要发费5万元来提升管理水平和建立管理系统计算,市场潜在规模高达3万多亿。

参考关联项目

Big Data Landscape

阿里云 MapCompute

国云大数据魔镜

四、项目方案

1. 上层应用模块的功能设计方案



● 销售管理模块

集成了客户关系管理、销售订单管理和分销管理的功能,使营销数据更加完整,并且加入营销数据分析,帮助销售人员提高销售业绩、增加客户的粘附度、提高客户的终身价值。

● 生产管理模块

本模块的功能目标是实现智能排单和自动生成生产计划和物料需求计划,提升了生产计划的准确度,有效的降低库存成本,保持产供销平衡,提高预测和决策能力。

● 人力资源管理模块

本模块通过对人力资源数据的分析,建立企业的目标、实现智能的绩效管理和薪酬管理。

企业员工可以根据公司的目标制定工作计划:

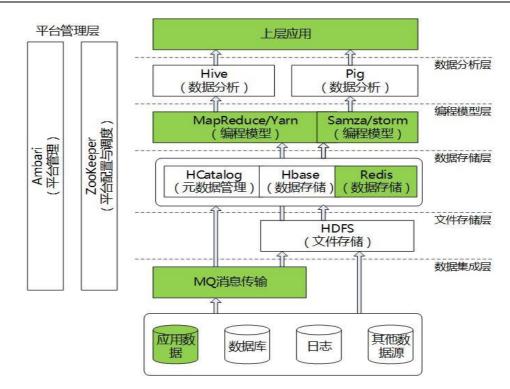
系统自动跟踪工作计划的实施情况:

系统对计划实施的情况进行反馈, 员工对实施情况进行检查;

系统对的实施结果进行评估, 员工根据评估优化和改善。

● 办公自动化模块

本模块集成了消息通知、邮件、即时通讯、办公管理的功能。通过对办 公数据的分析,得出智能提示和智能通知功能,大大节约人力成本,提 高企业的工作效率。



2.核心架构层方案

项目核心架构可以分为5层:

数据分析层:使用 Hive 搭建企业信息的数据仓库, Pig 进行这些海量数据数据的分析。

编程模型层:采用 MapReduce 分布式并行编程模型以及 YARN 进行资源的管理和调度,并使用 Samza 和 Storm 进行数据流的实时计算企业的的大规模业务信息。

数据存储层:利用 HcataLog 记录数据元信息,利用分布式数据库 Hbase 存储海量数据,并用高速缓存 Redis 进行数据信息的缓存,提升数据读写的性能,给我们的用户以更完美的体验。

文件存储层:对于企业数据分析来说,数据的可靠性,一致性,存储的容灾性都是直观重要的,我们采用的分布式的文件系统 HDFS,利用多点保存以及合理的数据一致性算法。保证了数据存储的可靠性,容灾性。

数据集成层:这一层集成了大规模的企业源信息数据,我们利用 MQ 消息传输记录这些数据元数据信息,并将这些数据存储与分布式文件系统 HDFS 之上。

五、项目经费

预算支出科目	支出金额(元)	预算根据及理由
租用云服务器	3000	项目需要购买多个云计算资源
图书资料	1000	技术方案的部署需要参考大量数据资料

备注:项目经费应专款专用,经费开支必须严格按照学校的财务制度执行,经费的开支范围包括①元器件和原材料(包括实验耗材、药品、化学试剂等);②加工制作与试验费;③图书资料;④学术交流(附参与学术会议的会议通知原件);⑤调研费用(经指导教师和学院审核,确需外出调研,包括来往车票、住宿费,应提交指导教师签字的说明材料和相关通知文件);⑥在学术期刊发表论文的版面费(须提供杂志社的录稿通知书原件);⑦专利申请费(须提供国家知识产权局的专利受理通知书原件);⑧其他与项目研究有关的费用。

六、预期成果

2017年5月2018年3月,完成项目的设计方案开发,并进行大规模测试。2018年3-5月,完成项目上线的实际应用。

导师意见:

签 名:

年 月 日

学院评审小组意见:

组长签名:

年 月 日

学院(系)意见:				
	签名盖章:			
		年	月	日
学校意见:				
	(盖章)			
	年 月 日			