文章编号: 2096-1472(2018)-01-35-03

DOI:10.19644/j.cnki.issn2096-1472.2018.01.010

# 基于网络数据包的大数据分析程序设计与开发

#### 叶 娟

(广东科学技术职业学院广州学院, 广东 广州 510640)

摘 要:基于网络数据包环境下大数据分析任务对精准度与速度要求不断提升,本文在此基础上重点论述大数据处理开展的理论环境。基于传统技术方法中存在的弊端,重点论述技术方法优化解决措施,提出大数据分析程序的有效设计方法,从而达到最佳数据处理效果,帮助提升网络数据包文件下载速率。

关键词: 网络数据包, 大数据分析, 程序设计中图分类号: TP311 文献标识码: A

## Design and Development of Big Data Analysis Based on Network Packets

YE Juan

(Guangzhou School, Guangdong Polytechnic of Science Technology, Guangzhou 510640, China)

Abstract: Given that the requirements for accuracy and speed in big data analysis is in rapid escalation in the network data packet environment, this paper mainly discusses the theoretical environment for big data processing as well as technical optimization solutions based on the existing drawbacks in traditional technology and methods and proposes an effective design method of big data analysis program, aiming to achieve optimum results of data processing and to help accelerating the download speed of network packet files.

Keywords:network packets;big data analysis;program design

#### 1 引言(Introduction)

网络数据包技术下,数据处理可以同时完成更大任务量,对于发展期间比较常见的数据处理隐患问题,通过数据分析处理也得到了有效预防控制。大数据分析程序设计中,充分利用云计算技术来提升数据处理速度,将数据处理任务结合环境来进行,在处理数据的同时进行数据对接传输,满足程序运行控制需求。面临大工作任务环境下的数据处理,程序内会自动将处理能力从TB级向FB级转变,最大程度优化数据处理系统应用效率、避免在最终功能上因此受到影响。批量处理符合大数据程序运行特征。面对大数据环境所开展的批量处理,将客户信息建立成为一个整体,共同进行数据处理。根据所得结果进入到更细致的分析环节,从而实现程序运行的数据处理任务[1]。

MapReduce属于当前常用的大数据处理系统,在处理过程中能够满足大规模群任务需求。在该系统中,将不同任务需要之间整合在一起,共同形成大数据处理模型。该项处理技术属于主流技术,应用后自动化性能更理想,处理运算能力强大,同时操作界面更简单且利于理解<sup>[2]</sup>。该项技术在数据处理期间可以将复杂的处理过程隐藏起来,节省显示界面所占据的内容空间。同时通过学习记忆功能来实现更深层次的数据挖掘和算法应用,为管理计划进行建立一个适合的基础环境,记忆所处理的数据并编排成为历史数据库。面对海量数据处理任务,首先会进行历史数据对比,参照记忆功能来

处理剩余部分数据,自然能够节省大量时间,这也是该项技术最常使用的原因。随着计算机网络技术不断发展,也产生了更多的大数据处理系统,可以满足不同类型的高端数据处理需求。本文将针对这一技术性方案展开探讨。

# 2 传统网络数据包大数据分析模式(Big data analysis mode for traditional network packets)

#### 2.1 BSP计算模式

数据信息处理采用周期性方法,在处理系统中将数据整体划分成为多个单元格,再依次完成各个单元格的处理任务,在精准度上可以达到大数据分析标准。但由于划分单元格过多,面对海量数据处理分析需要一段时间,在时间和速度上难以达到最佳使用效果。每个数据均会在单元格内完成分析计算,对于计算期间的各项处理分析任务,机遇网络数据包环境下面临大量处理任务,这也是当前常用技术性方法,其数据分析构成模型如图1所示。

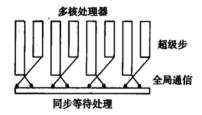


图1 BSP模型状态 Fig.1 BSP model state

图1表示为BSP计算模式的结构状态,在系统中会自动将需要分析处理的数据进行汇总,整理成为等待处理的数据,划分单元结构基础上共同进行多核处理,通过这种方法来帮助节省时间。在超级处理步骤中会同时启动多项处理器,共同参与到全局通信系统中,将网络环境中所获取的数据通过这种方法来整合成为需要的模式。传统处理模式中需要大量时间来调动各个处理器,在效率与时间上急需调整。

#### 2.2 BSP模式中的缺陷

该模式中存在一个超级步骤,在此环节中应用处理器来对数据进行处理、传输都会浪费大量时间,同时面对多项信息处理任务时在信息容量上也很容易达到上限,数据开发期间更是难以达到最佳使用效果。基于网络数据环境下实现大数据分析程序开发设计,需要一个适合的数据存储环境,而在BSP模式下并没有达到这一效果,处理问题的能力自然也因此受到影响<sup>[3]</sup>。超级处理步骤不仅在存储空间中要求大,数据传输也会同时占据大量网络空间,导致企业信息传输任务进行遇到网络延迟,难以在预期时间内完成分析任务。BSP模式中存在的主要弊端是处理量与网络需求之间的矛盾。该种开发模式在运算方式上需要继续优化改变,这样才能达到最佳效果。对于运算期间可能会产生的问题更要采取预防控制中方案,以免最终的功能稳定性因此受到影响。

## 3 基于网络数据包的大数据分析程序设计构想 (Design conception of big data analysis based on network packets)

#### 3.1 SAT模型逻辑分层

逻辑层共由三个结构组成。首先S层是针对数据进行分隔,避免分析处理过程中数据之间出现干扰。A层中表示数据分析处理和处理过程中的任务操作稳定性,会根据大数据处理特征来有针对性地规划相关信息,从而达到最佳处理分析效果。T层中则表示数据分析过程中的加工整合和具体包装情况,与实际情况之间相互对应进行,可以根据不同数据分析任务中的特征来进行包装优化,符合网络数据包大信息量使用特征,最终控制任务进行效率也更高速。SAT模型的具体大数据分析流程图如图2所示。

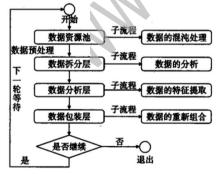


图2 SAT模型大数据分析流程

Fig. 2 Big data analysis process of SAT model

由图2可知,在SAT模式中,大数据处理各个任务划分十分严格。对于分析处理过程中可能会产生的问题,在此环境下也能得到充分解决。有关于现场任务处理中可能会产生的问题,各个模块中可以加入预防程序,在大数据处理过程

中启动这一程序,从而达到最佳分析效果。与之相匹配的数据分析处理任务,最终结果也更加理想。为提升大数据分析速度,五个功能模块采用统一系统进行总控制,最大程度减少中间处理环节,这样在处理时间上更优化节约。该模式下可以充分调动多个系统之间相互合作,共同完成处理分析任务,从而实现对控制计划中的综合控制,分析层之的间转换也更高效和谐,分析结果也能直接通过程序运行展现出来。

#### 3.2 SAT大数据分析程序实现流程

#### (1)数据预处理

在使用需求的基础上对数据信息进行预处理,并结合实 际情况观察所分析数据是否在使用中与实际情况可以保持一 致。在网络数据包的基础上所开展的数据分析和处理,需要 充分利用云计算技术来帮助处理数据。预处理后能够将分析 任务进行模块划分,确定其具体属性所在,在此基础上帮助 提升最终功能,并结合各项技术性方法来探究相关问题的解 决处理措施,建立起各个数据库预处理联系性,以便更好地 确定其属性[4]。模糊预处理可帮助节省大量时间,并筛选出其 中重要信息内容, 进行接下来更深入的任务分析, 发现问题 后充分探讨其解决规划措施,从而达到最佳处理效果。模糊 处理后筛选出的数据需要进行二次精准分析,进入到分析程 序中的循环部分[5]。因此在对程序进行构建期间,循环流程也 是不可缺少的。将多项任务相互结合共同高效运行,可以达 到理想的使用效果,并为管理任务开展建立一个适合的基础 环境。设计筛选后的数据集合,二次循环过程中的数据分析 可以针对这部分信息来进行,从而帮助节省大量时间,进入 到系统分析平台中。

#### (2)淘汰失效数据

预处理和分析任务结束后,都需要进行现场预处理,观察到信息结果与实际情况之间存在差异性联系后,将无用数据进行淘汰处理。系统程序内自动完成这一任务,将失效数据删除后可节省大量存储空间,数据分析任务进行也可以在特征数据库中进行。为了解决系统复杂性带来的挑战,人们需要结合大数据的价值稀疏性和访问弱局部性的特点,针对能效优化的大数据分布存储和处理的系统架构,以大数据感知、存储与计算融合为大数据的计算准则,在性能评价体系、分布式系统架构、流式数据计算框架、在线数据处理方法等方面展开基础性研究,并对作为重要验证工具的基准测试程序及系统性能预测方法进行研究,通过设计、实现与验证的迭代完善,最终实现大数据计算系统的数据获取高吞吐、数据存储低能耗和数据计算高效率,建立如下的程序文件代码,将其应用在程序中,控制各个系统模块运行。

parent:[parent\_task\_name]#父任务,将有用历史数据直接继承下来,帮助节省信息比较分析所用时间。

task\_name:[task\_name]#程序系统内所分析数据文件的 具体名称,将其与系统内控设计文件相互结合起来。

[phase\_name]: #任务执行的程序代码。

-step:[plugin\_name]。[module\_name]#程序系统运行中的插件设计。

[attr\_name]:[attr\_value]

-step:[plugin\_name]。[module\_name]
[phase\_name]:

-step:

筛选程序设计完成,在以上程序中能够实现不通区域之间跨越分析,适用于大数据环境,对于不同环境分析使用需求,改变程序内部控制模式,可以实现与数据环境符合的分析模式。网络数据包获取速度受网络环境稳定性直接影响,如果不能控制好当前需要的环境和各项控制计划,最终的工作稳定性也将会因此受到影响<sup>[6]</sup>。程序系统开发设计中考虑网络平台信息传输能力,在此基础上完成程序系统更高效的开发任务。

#### (3)基于网络数据包的文件下载

系统内会根据大数据分析任务来进行相关文件包下载流程如图3所示,并将其应用在适合的控制模块中。执行程序后首先进入检索环节,将其连接到文件夹有用价值判断中。判断该文件夹无用,则进行文件夹重新获取,直到所确定的文件夹范围与使用需求之间保持一致,进行到下载环节中。基于网络数据包基础上下载正确的文件夹,才能进入到更理想的数据使用环节中,接下来的大数据分析计划也能与之保持一致。

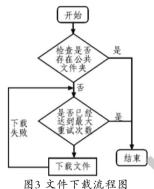


Fig.3 Flow chart for file downloading

当检测所获取的文件夹能够与使用需求之间保持一致,则会直接进入到大数据分析环节中。数据分析过程中比较容易出现隐患,要加强文件夹现在阶段的审核,确定数据精准程度后再次进入到分析环节中,更精准高效地完成任务。

#### (4)程序系统功能初始化

初始化功能能够对所分析的大数据文件进行恢复,节省 大量信息筛选所用时间,并直接将系统运行中需要的信息引 人其中,初始化流程框图如图4所示。

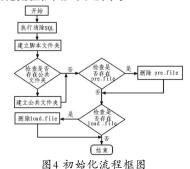


Fig.4 Flow chart for initialization

初始化具有清除和恢复功能,程序开发中设计多项控制功能,并将其规划到具体方案中,落实完善后才能达到更理想的程序使用效果,初始化过程中会自动恢复系统不需要的信息内容,以便节省处理空间,为所开展的各项设计任务建立一个适合的基础。检查过程也是一项筛选功能,判断所存在的问题,并加以优化解决。文件夹处理分析过程中程序内各个模块之间相互配合运行,共同参与完成信息检索任务。

#### 4 结论(Conclusion)

互联网、物联网、云计算技术的快速发展,各类应用的 层出不穷引发了数据规模的爆炸式增长,使数据渗透到了当 今每一个行业和业务领域,成为重要的生产因素。大数据因 此成为社会各界关注的新焦点,大数据时代已然来临。为了 应对不同的业务需求,以Google、Facebook、Linkedin、 Microsoft等为代表的互联网企业近几年推出了各种大数据处 理系统,深度学习、知识计算、可视化等大数据分析技术也 得到迅速发展,已被广泛应用于不同的行业和领域。本文根 据处理形式的不同,介绍了批量处理数据、流式处理数据、 交互处理数据和图数据四种不同形式数据的突出特征和各自 的典型应用场景,以及相应的代表性处理系统,并总结出引 擎专用化、平台多样化、计算实时化是当前大数据处理系统 的三大发展趋势。随后,对系统支撑下的深度学习、知识计 算、社会计算与可视化四类大数据分析技术和应用进行了简 要综述,总结了各种技术在大数据分析理解过程中的关键作 用,即深度学习提高精度,知识计算挖掘深度,社会计算促 进认知,强可视化辅助决策。

#### 参考文献(References)

- [1] 周情涛,何军,胡昭华.基于GPU的Spark大数据技术在实验室的开发应用[J].实验室研究与探索,2017,36(1):112-116.
- [2] 程平,张砾.大数据时代基于云会计的房地产开发企业税务管理探析——以A集团为例[]].会计之友,2017(8):134-137.
- [3] 王岗.基于大数据的用户体验分析应用——"互联网+"环境下新品研发质量管理探索[[].上海质量,2017(6):26-27.
- [4] 黄冬梅.大数据时代下地勘业务转型发展的实践与思考——以五矿勘查开发有限公司实践为例[J].中国国土资源经济,2017,30(3):9-13.
- [5] 孙秀春,廉新宇.基于大数据分析的高职院校物联网专业创新创业教育模式研究与实践[J].人才资源开发,2017(4):234-234
- [6] 吕慧."95后"大学生思想变化特点和规律研究——基于大数据思想分析维度[]].人才资源开发,2017(8):138-140.

#### 作者简介:

叶 娟(1973-), 女, 本科, 副教授.研究领域: 计算机应用.