



设计院大数据挖掘系统的建设思路研究

Research on construction ideas of institute big data mining system

董春辉,徐瑞/DONG Chunhui, XU Rui

中国移动通信集团设计院有限公司山东分公司 济南 250101

Shandong Branch, China Mobile Group Design Institute Co., Ltd., Jinan 250101, China

摘要:随着建设项目及新技术的增多,设计院掌握的相关资料数据也越来越多,这些资料数据不是孤立存在的,且相互之间的关联性越来越大。面对海量资料数据,设计人员无法快速、全面地检索定位,需要建设一套大数据挖掘系统来帮助完成。从系统架构及功能实现等方面,为移动设计院的大数据挖掘分析系统建设提供一种思路。

关键词:大数据;数据挖掘;分析系统

Abstract: With the increasing of construction projects and new technologies, the institute grasped more and more relevant information data. But these information and data didn't exist in isolation, the correlation between them was also growing. Faced with massive information and data, designers couldn't retrieve positioning quickly and comprehensively, and need to build a big data mining system to help complete. From the aspects of the system architecture and function realization, an idea for big data mining analysis system construction of mobile design institute was provided.

Key words: big data, data mining, analysis system

1 引言

随着建设项目及新技术的增多,设计院掌握的相关资料数据也越来越多,而这些资料数据并不是孤立存在的,相互之间的关联性也越来越大。因此,设计院需要一个大数据挖掘分析平台来挖掘、关联和分析数据、资料的相关性,以便更好地为客户服务。

2 系统架构

考虑到移动设计院业务工作多为外出勘察等移动办公,故该系统设计为客户端、中心系统以及数据源采集3层体系架构。系统主要功能及架构如图1所示,功能介绍如下。

①客户端主要用于实现勘察数据、沿途无线信号覆盖类数据的上传以及面向最终用户的查询工作。

②中心系统包括平台层、处理层以及管理与展示层,分别用于系统的承载、数据的采集、建模、挖掘分析以及管理呈现等工作。

③数据源采集主要包括内/外部数据的导入以及客户端数据的上传等工作。

3 系统功能思路

系统面向不同类型的用户提供不同的功能,介绍如下。

①设计人员:提出各类分析需求,并使用系统提供的应用平台进行实际分析,汇总分析报告等。

②数据分析人员:根据设计人员提出的分析需求,寻求相应的数据分析实现方法,在系统中进行相应的分析规则和数据关联关系配置及维护,还有联机分析和数据挖掘算法的实现和维护。

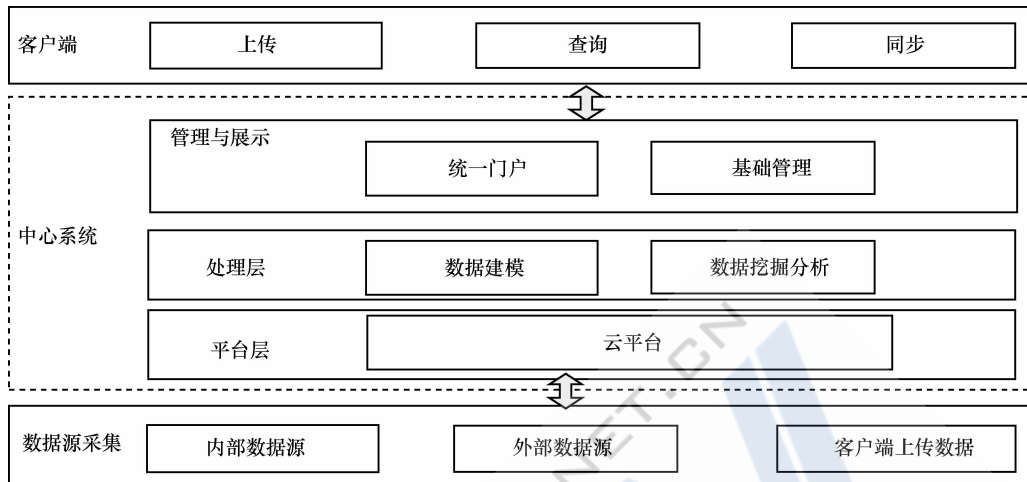


图1 大数据挖掘分析系统架构

③运维人员:进行系统自监控、业务分析告警、云平台、外部系统接口数据的运维管理等。

大数据挖掘分析系统各功能模块详细说明如下。

3.1 数据源采集

包括对内/外部数据和客户端的采集以及自动上传等功能。

3.1.1 内部数据源

内部数据源为移动设计院各系统的建设方案资料、无线覆盖建设类资料、集团客户需求类资料等。

(1)各系统方案资料

包括各类系统的建设依据、建设规模、功能需求、接口需求、软/硬件测算及配置方案、设备信息以及网络部署、平面部署等对应资源占用。

(2)无线

包括城区/县/乡覆盖方案、站址站距、经纬度、下倾角、设备信息、配套设备信息、机房土建信息以及勘察资料。

(3)集团客户

包括集团客户需求、实现方案、资源使用情况等。

3.1.2 外部数据源

外部数据源为从移动设计院外部可以得到的数据,包括兄弟单位/权威机构以及互联网。

(1)兄弟单位/权威机构

包括各种知识库、各级别规范以及发表的各类文章。

(2)互联网

对技术网站上的观点、思路、方案、舆情以及新

业务需求等进行收集。

3.1.3 客户端上传数据

主要为勘察资料、沿途无线信号覆盖情况、查询需求、查询记录及手机/主机客户端间同步数据等。

3.2 中心系统

中心系统又分为平台层、处理层以及管理与展示层。

3.2.1 平台层

(1)云资源

通过虚拟化技术,将主机和数据库等物理IT资源虚拟化成虚拟机、虚拟存储,并通过云管理手段将虚拟机、虚拟存储整合成宿主机、资源池等云资源。

(2)云资源管理

实现对云资源的统一分配和动态调度,包括对计算资源和存储资源的统一管理,并支持对虚拟机的创建、修改、迁移、克隆、回收等全生命周期管理。

3.2.2 处理层

(1)数据建模

依据现有业务及掌握的数据进行数据建模。

(2)数据挖掘分析

①ETL处理:将获取的数据进行数据抽取、清洗、转换,按预先定义好的数据仓库模型,将数据加载到数据仓库。

②数据分析:采用数据汇总、联机分析和数据挖掘等技术手段,对ETL处理后的原始数据进行提炼和精加工,形成指标、多维数据和数据挖掘结果等各类数据。

③数据存储:实现对数据仓库的集中存储和管理,并根据需求建立面向业务分析的数据集市。数据存储的内容包括经 ETL 处理后形成的原始数据和经数据分析后的数据。

3.2.3 管理与展示层

为设计人员提供各层次数据分析功能和分析成果的可视、可追溯展示功能;为后台工作人员提供各类管理维护功能。

(1)统一门户

①数据查询:为设计人员提供数据查询功能,包括即席查询、OLAP 查询、统计分析、预测评估和历史查询记录功能。

②分析报表呈现:根据数据分析结果,定时产生或即时查询日常分析报表和专题分析报表,报表格式通过报表模板进行管理,可灵活配置。

(2)基础管理

①配置管理:支持由数据分析人员配置 ETL 处理后的各类型数据之间的关联关系,配置数据分析(数据汇总、联机分析、数据挖掘)的参数、规则、算法等,同时由维护人员配置固化的分析,出现分析结果异常的告警阈值。

②指标管理:将 ETL 处理后的指标数据按定义的指标体系进行统一配置和管理。

③权限管理:包括用户管理、权限角色管理。系统前台权限分为页面权限、数据权限、权限控制的粒度到页面以及页面展示的内容,管理员可以将权限分配给角色,然后将角色分配给用户。

④系统自监控:监控本系统的运行情况,包括对物理资源(主机、数据库)、云资源、应用进程运行情况的监控。

3.3 客户端

客户端的主要功能为及时数据、勘察资料的上传以及前台人员自定义查询等。

3.3.1 上传

包括准实时上传及主动上传两类功能。

(1)准实时上传

准实时上传内容主要包括无线信号、Cell ID 等覆盖情况以及查询记录,可将详细信息缓存至本地,并在业务闲时压缩上传。

(2)主动上传

主要将勘察数据、照片等上传至中心系统。

3.3.2 查询

主要用于前台人员自定义查询以及查询结果的呈现,并记录查询信息与浏览信息。

3.3.3 同步

用于不同类型客户端间的数据同步,即手机客户端与 PC 客户端间的数据同步。

4 系统建设思路

4.1 数据源采集

内部数据可与现有工单类系统对接,将产品成果直接上传。

外部数据可采用爬虫类技术实现基于关键字的数据采集。

4.2 中心系统

4.2.1 平台层面

可借助现有云计算平台资源池或使用 VMware 构建资源池,采用 Hadoop 实现并行计算。

4.2.2 数据建模和分析

在建设初期,基于文本字段的对比和关联呈现,如联系互联网中舆情类信息补充业务功能需求,新建系统的业务需求是否在现有系统中已经实现,可避免重复建设,在一定规模下的软/硬件实际配置用以修正能力测算,集团客户的需求在哪些项目中采用了哪种方案实现。

在建设中期,建设基于 CAD、Visio 等工具类软件的深入挖掘分析,可以更深层次地实现对现网资源占用情况的统计分析等。

在建设后期,实现基于图片、视频类信息的关联分析等。

4.2.3 系统管理

关键为系统权限设置,其原则为依据不同项目负责人等级划分来分配不同的在线阅读权限,如国家级项目负责人可在线查询浏览全国项目及各省项目的相关资料,省级项目负责人可在线查询浏览本省内项目的相关资料。

对于越权类查询浏览、上传及下载类操作,均需向上级领导报批,并由系统记录相关日志。

4.3 客户端

关键为除了记录主小区信息外,同样记录邻小区相关信息。

4.4 实现目标

(1)以传统新建核心网/IT类项目为例

①设计人员接到任务,将用户需求通过客户端输入系统进行关联查询。

②系统将查询后台对应技术资料文档,可反馈给客户端是否有类似的功能已在其他系统中进行了建设。

③可查询是否有相关规范对此类需求提出了建设要求。

④如果存在类似功能,则视权限情况推送给设计人员相关文档在线预览,并列出占用现网资源情况,同时根据以往的查询浏览记录,提示还有哪些资料可能相关。

⑤有针对性地列出有哪些舆情类信息能反映出需要加强哪部分的建设,并提示是否有新趋势、新技术与新观点,以提供给设计人员新角度与思路。

(2)以传统无线基站建设项目为例

①设计人员到达目的地后,输入经纬度、高度、楼房类型等现场信息进行相关性查询。

②系统将查询后台,反馈给设计人员类似站点布置、承重加固、使用抱杆等建设方案参考。

③系统将提示本期工程项目的建站原则以及站址规划表,并可根据相关要求和算法自动算出偏离度。

④系统将相关性信息进行提示,如周边信号覆盖情况,从互联网舆情中是否可发现,此处会有大型活动计划等,供设计人员参考。

5 结束语

本文为设计院大数据挖掘分析系统建设提供

了一种建设思路,供类似工程咨询单位系统建设参考。该系统的建设和使用可以使设计人员提高效率、避免冲突、减少思维盲点、增加新思路。

参考文献:

- [1] VIKTOR M S, KENNETH C. Big Data: a Revolution that Will Transform How We Live, Work and Think[M]. Zhejiang People's Publishing House, 2012.
- [2] FRANKS B. Taming the Big Data Tidal Wave[M]. Posts&Telecom Press, 2013.
- [3] HAN J, KAMBER M. Data Mining: Concept and Techniques[M]. China Machine Press, 2012.
- [4] ANAND R, JEFFREY D U. Mining of Massive Datasets[M]. Cambridge University Press, 2011.
- [5] 王星. 大数据分析:方法与应用[M]. 北京:清华大学出版社, 2013.

作者简介



董春辉(1980-),男,中国移动通信集团设计院有限公司山东分公司高级工程师,多年从事数据及IT支撑系统咨询设计工作。



徐瑞(1981-),女,现就职于中国移动通信集团设计院有限公司山东分公司,多年从事无线网络咨询设计工作。

· 30日扫描 ·

PP助手2016战略:利用阿里矩阵发力移动分发

阿里移动近日宣布,为加强移动应用分发,2016年PP助手将启动“数据+”战略。具体来说,阿里移动将通过阿里云、UC浏览器等多个阿里系产品,协助PP助手实现全景式精准分发,此外还将上线针对开发者的移动应用广告平台。

阿里巴巴移动事业群PP助手总经理张博表示:“移动互联网用户正变得越来越成熟,需求进

一步细分化,移动应用分发和用户需求产生错位的现象越来越常见,反映在用户的行为特征上就是自主性逐渐增强,使用应用商店的生命周期开始缩短。在这种形势下,要同时满足用户和开发者的需求,应用商店必须做到个性化精准分发,对此,PP助手结合自身可利用的优势提出并践行‘数据+’战略。”