

大数据技术在智慧机场领域的应用研究

顾 凯

(东软集团(上海)有限公司, 上海 200241)

✉guk@neusoft.com



摘 要: 随着信息技术的发展, 大数据技术在公共服务中的交通、医疗、教育、预测服务等领域得到广泛应用。本文从大数据存储、挖掘、分析和应用等几个角度重点介绍了机场大数据分析平台的构建方法, 在机场转型升级方面提供了信息整合及数据分析模式的创新, 以大数据技术在智慧机场领域的应用为研究内容, 通过多个案例讨论大数据技术在智慧机场中的应用, 从而加深人们对智慧机场的认识。

关键词: 大数据; 智慧机场; 数据决策; 机场运营

中图分类号: TP391 **文献标识码:** A

Application of Big Data Technology in the Field of Smart Airports

GU Kai

(Neusoft (Shanghai) CO., LTD., Shanghai 200241, China)

✉guk@neusoft.com

Abstract: With the development of information technology, big data technology has been widely used in public services such as transportation, medical care, education, and forecasting services. This paper proposes a method of constructing big data analysis platform for airports from several perspectives such as big data storage, mining, analysis and application. In terms of airport transformation and upgrading, it provides innovations in information integration and data analysis models. It takes the application of big data technology in the field of smart airports as the research content, and discusses the application of big data technology in smart airports through various cases, so that people may have a deeper understanding for smart airports.

Keywords: big data; smart airport; data decision-making; airport operation

1 引言(Introduction)

自2019年民航局正式印发《推进四型机场建设行动纲要(2020—2035年)》(以下简称《纲要》)以来, 国家加快了四型机场建设的步伐。《纲要》明确, 到2035年将实现标杆机场引领世界机场发展, 全面建成安全高效、绿色环保、智慧便捷、和谐美好的四型机场, 为全方位建设民航强国提供重要支撑^[1-2]。

在智慧机场推动转型升级, 加快信息基础设施建设、实现数字化, 推进数据共享与协同、实现网络化, 推进数据融合应用、实现智能化, 切实保障信息安全的主要建设

过程中^[3], 大数据技术是数据共享与协同、数据融合应用的重要研究内容。

基于互联模式设计万物互联的机场数据中心平台架构, 研制专用的大数据交互链路, 实现采集、交换、数据比对、数据处理、数据服务和数据应用全周期建设, 建立统一、灵活、高效、便捷的机场数据中心是主要目标。为此, 基于从2017年带领团队参与华东某机场大数据治理及相关应用工作中积累的经验, 本文提出了机场大数据应用新模式的探索, 探讨机场全局数据共享和分析, 打造机场协同决策系统和运行数据库, 实现运行、安全和服务可视化运营解决方案。

效传输大量数据的工具，提供了多源异构信息融合的功能，可以将外部多源异构数据库中的数据导入HDFS等相关系统中，或者从内部系统中导出数据到外部多源异构的关系型数据库和数据仓库。

(6)分布式协议服务：提供分布式应用程序的协调服务，实现系统命名服务和同步服务等功能，使得整个分布式的大数据商业智能与数据挖掘分析软件内部模块可以步调一致地工作。

5.1 数据存储

提供大数据平台部署工具，提供组件部署管理及动态增加机器节点管理工具；提供组件的运行状态、组件的负载情况监控及组件的启动、停止、移除等管理。当组件故障自动迁移，节点组件出现故障时，集群中其他节点中的相应组件自动接管故障组件的工作，保证组件正常服务。支持分布式存储模块，提供分布式文件的存储、存储副本的管理；分布式计算模块提供分布式数据计算处理及数据的分布式均衡访问。支持数据采集与预处理模块，提供数据采集与预处理框架，实现数据向大数据平台汇聚的功能。

5.2 数据仓库DWH

数据仓库DWH是一种建立在Hadoop之上的数据仓库架构，是为实现海量数据的高性能处理而设计的，可以高效地对TB甚至PB级的数据进行处理。

DWH在设计上是一种批处理系统，它定义了一种简单的类SQL查询语言NoSQL，让熟悉SQL的用户可以非常容易地进行查询。

DWH自身并不存储用户数据，而是通过接口访问用户数据，这就使得DWH支持各种数据源和数据格式，可以与其他大数据的应用兼容。

5.3 数据挖掘

数据挖掘模块是具有可拓展算法的机器学习框架，能够快速创建智能应用，可以帮助系统智能地、准确地识别知识。

数据挖掘模块内嵌了一些基于使用场景的算法，例如：

(1)推荐引擎算法：通过分析用户的行为历史记录来推算用户最可能关注的事物或习惯。

(2)聚类算法：通过分析将一系列相关的事物划分为相关性相近的群组。

(3)分类算法：通过分析一组已经分类的事物，将其他未分类的事物按同样规则归入相应的分类。

(4)相关事物分析算法：识别出一系列经常一起出现的事物组。

数据挖掘模块所处理的场景，经常是伴随着海量用户使用数据的情况。通过将算法的输入、输出和中间结果构建于HDFS分布式文件系统之上，而具有高吞吐、高并发、高可靠性的特点。最终，使业务系统可以高效快速地、实时地得到分析结果。

5.4 数据分析

数据分析模块是一个大规模数据分析的平台，它为海量数据的并行处理提供了操作以及编程实现的接口。

数据分析模块包括可视化分析工具、流程建模工具、文本信息抽取工具、统计分析工具、算法开发支持工具。可视化分析工具提供影像、拓扑关系等数据的快速展示，提供基本的可视化挖掘功能；流程建模工具提供业务处理流程的建模、运行和监控，实现快速数据分析原型建立；文本信息抽取工具提供了命名实体识别、文本语义提取等功能，是文本结构化分析的基础；统计分析工具提供了参数统计、统计检验等分析功能；算法开发支持工具提供了算法开发、测试、集成的环境，实现分析算法的快速开发和应用。

5.5 ETL

ETL(Extract-Transform-Load,数据抽取、转换、装载的过程)功能主要有如下几点：

(1)数据抽取功能

支持主流数据接口，如Oracle系列数据库、SQL Server、Sybase ASE、DB2、MySQL等，并实现各类数据库接入模式，简便易操作，支持的数据类型全面。

(2)数据清洗功能

具有各种数据检测函数，能够帮助用户发现异常数据。在数据清洗过程中，支持对数据进行预览。具有多种数据转换函数，可以帮助用户修正异常数据。数据转换异常时支持数据自动转存，帮助用户分析数据情况。

(3)数据转换功能

支持各种字段级的映射转换，如类型转换、字段运算、参照转换、字符串处理、空值处理、日期转换、聚集运算、既定取值、字段切分、字段合并等；支持已抽取记录的切分、过滤；支持表的动态创建和源数据表模式到目标表模式的自动映射；支持数据库存储过程及函数的调用；提供了大量的API函数，可以满足各种数据转换需求；支持脚本编程，可相当灵活地进行数据转换规则设计。

(4)数据装载功能

支持将源数据库中的记录分发到不同的目标库，针对每个分发目标库都可以设置分发的条件，同时支持批量装载。

(5)数据交换功能

数据交换功能提供三种数据交换方式，包括：

①采用“数据交换接口+前置机+数据交换软件”进行实时交换。

②采用“人工抽取数据+前置机+数据交换软件”进行定时交换。

③采用人工报盘方式进行定时交换。

(6)规则设计功能

规则设计功能为用户提供向导式的数据整合规则设置，按照类别、处理阶段、具体整合功能(动作)等，将设计过程

划分为阶段、步骤、动作。最小的调度单元是任务，任务之间可以有关联，手工调度则可以执行任何一个层次，最小到动作。

(7) 工作流调度功能

它是内置工作流建模工具和调度引擎，用户可自由定义任务的触发条件及执行路径，控制多任务的协同运行。支持多任务的并发运行，支持并发任务的同步完成；支持成功、失败、条件等三种路径，并支持基于全局变量的条件路径转移；支持任务的定时(可基于日、周、月、工作日、时、分、秒等)触发；支持系统命令、外部程序调用^[6]。

5.6 分布式协议服务

事务管理：在内部模块和组件之间创建事务，依据大数据的数据请求建立事务，包含事务监控、事务提交、事务回滚、事务创建等功能。

协议包管理：提供内部模块和组件之间协作的标准化协议管理和配置功能。

开放式的扩展接口：提高与外部大数据软件或其他软件之间的兼容、扩展和集成。

6 解决的问题(Key issues)

6.1 整合生产经营数据，强化协同，提供有效决策支撑

搭建统一的指标数据统计分析平台，整合集团生产经营源数据和合同管理，进一步强化各管理条线及业务条线的总体信息化水平，解决各下属部门数据收集烦琐、数据源复杂、人工干预多等问题，进一步提高数据完整性、准确性、及时性和一致性，通过生产经营分析平台对数据进行各类自动化统计分析处理，提升日常工作效率等，使机场及时了解生产经营业务实情，有效地帮助各项决策的制定。

6.2 预测航班延误时间

影响航班延误较为重要的因素有前序航班的延误情况、出发地以及目的地的天气状况、机场当前的出港率(平均每架飞机的起飞时间)和排队等候起飞的飞机数量、航路流量数据等。此外，还要考虑飞机加油、配餐、打扫卫生等多个因素，而且延误时间主要以实际起飞时间与计划起飞时间的差值为准，也就是说从飞机降落直至飞机起飞这段时间内的各个因素都会影响飞机起飞的延误状况。所以，在连接原始表的时候以飞机降落至起飞为主线构建数据表，以保证数据的全面性，并预测第二天的航班延误情况^[7]，如图2所示。

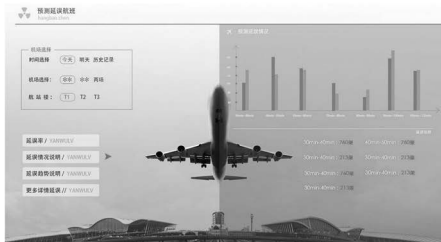


图2 大数据建模预测分析

Fig.2 Big data modeling predictive analysis

6.3 预测机场免税店畅销商品

影响机场免税店产品是否畅销的较为重要的因素有国际航班的客流量及其在机场的滞留人数、促销活动、新品活动、机场淡旺季等。由于POS系统中的商品分类为“单品、商品品牌、小分类、中分类、大分类”，单品和小分类的种类过多，中分类和大分类过于分散，不利于统计分析和预测，因此选择商品品牌作为商品的名称标签，以每天商品品牌的销售额度构建主线数据表^[8]。

7 结论(Conclusion)

大数据技术应用是机场大数据平台管理需求的集中体现，也是底层平台技术组件选择和数据源采集的指导。大数据技术既体现在传统核心业务和内部各信息系统的数据需求上，也体现在大数据挖掘所产生的数据潜在的价值上；既能够支撑管理及商业业务升级完善的功能，也具有一定的前沿性探索作用。

综合利用大数据技术底层数据资源和海量异构数据的处理能力，构建应用支撑层的分析工具和模型算法，以支撑未来机场的智慧化应用需求。

参考文献(References)

- [1] 陈燕,刘川民,曾珂.创新应用全覆盖站坪管理系统推进四型机场建设[J].航空航天科学与工程,2020(06):36-38.
- [2] 徐蒙.新时期智慧机场发展的思考[J].民航管理,2020(06):39-42.
- [3] 中国民用航空局.推进四型机场建设行动纲要(2020—2035年)[R].北京:中国民用航空局,2020:3-20.
- [4] 吉向东.基于大数据技术的智慧机场管理平台的设计与研究[J].信息系统工程,2020(07):40-41.
- [5] 魏君.民航“新”基建[J].大飞机,2020(06):60-65.
- [6] 翁业林,周泓,侯兵.面向企业级数据中心的分布式ETL研究与设计[J].软件工程,2018(12):15-18.
- [7] TIPAVINEE S, APICCHAT S. The impact of delay affecting airport efficiency: Sustainability perspective[J]. International Journal of Logistics Systems and Management, 2020, 37(4):445-464.
- [8] 徐敏.机场行业发展数字经济的思考[J].民航管理,2020(11):17-21.

作者简介:

顾凯(1975-),男,本科,软件工程师.研究领域:软件开发,信息研究.