

数据仓库与大数据融合的探讨

郭光福

(中国人民大学 信息学院, 北京 100872)

摘要 数据仓库与大数据技术的融合应用, 对提高数据提取、采集、存储、应用水平等有积极作用。因此, 针对不同业务的非结构化数据应用问题, 在对数据仓库与大数据融合过程进行分析中, 结合数据仓库建设的理论思想, 对大数据进行存储、处理、加工与分析, 并在大数据技术的视角下, 对数据库仓库优化进行调整, 旨在实现大数据与数据仓库的综合应用效果提升。

关键词 数据仓库 大数据 数据分析与整合

中图分类号: TP311

文献标识码: A

文章编号: 1007-0745(2021)11-0007-03

随着数据库技术的多元化发展, 其在各行业、各领域都有广泛的应用, 所以在业务管理系统中, 针对数据仓库与大数据之间的数据融合进行分析, 对进一步提高数据仓库的实际应用效果有积极作用。在数据库技术实现多样化发展中, 数据仓库是通过现有的数据库, 对数据信息进行抽取、加工以及使用, 并为管理工作的开展提供管理决策。在这一视角下, 数据仓库建设与发展应用数据库技术, 对现有数据信息进行管理, 满足业务管理与控制的综合发展需求。在利用数据库技术的基础上, 可通过数据挖掘与信息处理的方式, 实现数据管理与业务管理水平的综合提升。

1 数据仓库的概念

数据仓库并不是单纯的多数据库集合, 而是通过数据处理技术针对现有的数据库, 对数据信息进行提取、清洗、加工、处理以及使用, 并为业务管理以及管理决策等提供依据。在实现数据信息处理的过程中, 则需要通过大数据技术, 对数据仓库的数据进行分析与处理, 按照数据特征进行分类与处理, 从而提高数据信息的有效利用率。数据仓库与数据技术具有联系性, 利用数据库技术对数据信息进行处理, 其中包含云数据处理、数据挖掘、AI技术应用等, 在数据信息处理与调整的基础上, 可实现数据处理水平提升^[1]。大数据在实际应用中, 可以减少数据冗余, 并消除数据壁垒, 在数据挖掘以及智能处理技术的应用下, 对数据信息进行融合与处理, 从而实现数据采集与挖掘。大数据在对数据处理与分析时, 可以对大规模的、结构复杂的数据信息进行处理, 从而实现预测与决策分析水平的提升。从数据管理的角度进行分析, 数据仓库与大数据之间具有一定的关联性, 但是大数据对数据类型、数据处理等有积极作用, 在数据处理与信息控制的基础上, 可实现数据仓库的数据信息融合与控制。

对大数据的需求进行分析时, 则是在数据分析与数据管理的基础上, 实现数据预测与决策水平提升。从大数据与数据仓库融合的角度进行分析, 大数据与数据仓库之间比较相似, 但是数据量、数据类型、处理速度、数据结构

准确性等方面存在一定的差异性, 所以在利用数据仓库技术的过程中, 则是在预测大数据的基础上, 对数据仓库的数据处理过程进行完善, 从而满足数据统计与数据分析的综合需求^[2]。

2 大数据技术及其困境

在对大数据技术的实际应用进行分析中, 则是在非结构化数据库以及数据信息处理过程等方面进行分析, 在数据集合与数据挖掘处理的基础上, 可实现数据信息的综合处理效果提升。但是, 数据信息处理技术在实际应用中, 则需要对数据处理过程、数据处理参数等方面进行优化, 在实现数据集中处理与分析处理的基础上, 可满足数据仓库的数据处理与控制需求。但是, 大数据技术在实际应用中, 仍然存在以下几个难点。

1. 从技术层面进行分析, 大数据的研究与应用仍然处在初级阶段, 缺少完善、有效的数据解决方案, 对大数据信息处理以及数据挖掘处理等会产生直接的影响。在对大数据的实际应用进行分析中, 缺少体系化的模型, 虽然开源的 Hadoop 可以实际应用, 但是在数据信息处理的过程中, 数据处理技术的应用水平不足, 对数据应用以及数据仓库的数据挖掘等会产生直接的影响。在大数据信息处理与控制的过程中, 则可以通过数据挖掘与信息处理的方式, 实现数据挖掘与数据信息处理效果的综合提升^[3]。在建立完整化数据信息处理模型的基础上, 可通过数据处理以及信息评估的方式, 实现大数据处理效果提升。

2. 从大数据处理过程的角度进行分析, 在大数据技术的实际应用中, 大多是以 Hadoop 为框架, 但是在实际应用时, 其仍然存在非结构化数据, 在数据处理与数据分析中, 对数据仓库的数据挖掘与信息挖掘会产生直接的影响。通过数据仓库的应用, 在实现数据信息处理与控制的过程中, 数据仓库的处理过程仍然存在问题, 对数据信息处理以及数据管理等会产生直接的影响, 也会影响数据仓库与大数据技术融合的实际应用水平。而且, 在对结构数据进行优化的过程中, 仍然存在数据信息处理不到位的情况, 影响

大数据技术的实际应用效果与应用价值。

3. 数据安全问题。在大数据技术多样化应用的过程中,面对数据安全问题,对数据信息处理过程以及数据冗余等会产生直接影响。在 Hadoop、MPP 数据库、MySQL 数据流处理中,仍然存在数据库系统不完善的情况,对数据结构的安全性以及数据处理过程等会产生直接的影响。因此,在数据信息处理与数据开元处理的基础上,仍然需要对数据安全问题进行调整,从而实现大数据的数据安全控制水平提升^[4]。

4. 大数据应用与开发不足。大数据的实际应用与云计算过程相似,在对基础数据建设体系进行完善与优化的基础上,仍然需要从数据处理、数据维护以及数据分析等方面进行优化,在对数据处理过程进行优化的基础上,可通过数据仓库的信息处理与数据控制,解决大数据的操作与数据处理的综合问题。但是,在大数据处理的过程中,智能化、多元化的系统框架搭建,仍然存在数据仓库的开发问题,而且基础数据建设比较困难,后期的数据处理与资源维护等不足,对大数据技术的操作与控制等会产生直接的影响。

5. 大数据对业务管理决策的影响仍需进一步研究。大数据技术在实际应用中,对网络中的海量数据进行挖掘,可以发现数据中隐含的价值,但是在获取数据的过程中,仍然需要对每一个数据进行处理、加工以及分析,要对数据处理过程、数据分析准确性等方面进行探究,在数据分析与预测等方面仍有不足,对数据处理过程会产生直接的影响。在大数据信息处理与控制的过程中,数据存储与数据提取等不足,会影响数据挖掘与信息处理综合水平。大数据技术可以对数据仓库中的数据信息进行统计与分析,但是在数据信息处理与数据应用控制中,仍然存在数据挖掘与预测准确性不足等问题,对大数据应用会产生直接的影响。

3 数据仓库与大数据融合措施

3.1 数据仓库体系架构设计

为实现数据仓库的架构应用水平提升,在业务数据处理与管理的过程中,需要在数据仓库处理的基础上,对数据信息处理过程、数据分析等方面进行优化,可实现数据仓库的数据处理水平提升^[5]。在大数据技术需求分析的视角下,可通过数据管理与信息处理的方式,对数据仓库的不同特征数据进行提取与处理,从而实现数据应用效率的进一步提升。在加强对列式数据库内数据获取与分析的过程中,则需要对数据处理以及数据挖掘等方面进行优化,从而达到提高数据应用效率的目的。在对数据应用以及信息处理过程等方面进行分析时,则需要从业务数据源处理以及数据非关系型结构等方面进行处理,在实现数据融合的基础上,利用 Hadoop 对数据信息进行处理,从而实现数据仓库的数据分析处理效果提升。

从行业应用的角度进行分析,根据结构化程度共同作

用进行分析,则是在行业应用、决策分析的视角下,对分布式处理以及数据安全等方面进行处理,在数据分布与信息整合的基础上,可满足数据分析与技术处理的综合需求。大数据的架构在应用中,需要通过技术要素对行业数据、决策支持、分布式处理等方面进行优化,通过数据信息处理与业务整合,可满足数据仓库的数据挖掘处理需求。大数据架构在实际应用中,则可以对不同特征的数据信息处理过程、数据仓库的数据存储等方面进行完善,在数据结构化程度检验与分析的视角下,可满足数据分析与开发的综合需求。展示层是针对数据关系、请求响应、数据处理等方面进行优化,数据应用中,则是通过预测性建模与模拟,对数据源进行分析与处理,满足大数据挖掘与信息处理的综合需求。在对数据流进行处理的过程中,则是在建立分布式处理关系的基础上,对网络自动抓取的业务数据处理,并实现数据挖掘与信息处理的综合分析。

3.2 数据仓库与大数据的融合

从大数据需求分析的角度进行分析,在实现数据融合与信息处理的过程中,则需要在数据仓库的视角下,对业务数据关系以及数据处理过程等方面进行优化,在建立数据仓库的视角下,对数据提取过程、数据处理方式等方面进行优化,满足数据分析与处理的综合需求。因此,数据仓库与大数据在实现融合中,可实现数据的相互促进、补充与协作,将大数据技术下的数据仓库处理过程结合在一起,并对非结构化数据的采集、存储与传输等方面进行优化,从而实现数据信息的综合处理效果提升。非结构化数据信息的处理与控制,则是在数据仓库分析与信息处理的视角下,对 Hadoop 技术与数据仓库结合过程进行优化,并对数据采集、ETL、存储、处理、提取等方面进行综合分析,在数据信息处理与优化的基础上,将数据信息传输到数据仓库,从而实现数据处理与分析的目的。在建立数据仓库的基础上,可以在数据管理层通过数据仓库与 Hadoop 的结合,对数据处理方式进行优化,并通过数据分工来满足数据处理与数据分析的综合需求。利用数据仓库对非结构化数据进行处理,并对数据处理过程以及数据参数等方面进行调整,提高数据管理与信息处理的综合水平^[6]。

在创建 BI 层的基础上,可对数据存储的结构化数据库进行调整与优化,在实现数据信息处理与数据分解后,对不同数据进行分析与处理,对实现数据信息处理与数据分析水平提升等有积极作用。从技术层面进行分析,数据仓库与大数据技术在融合的视角下,其实际应用中,则可以通过数据处理与数据分析等方式,对数据驱动过程以及数据信息处理方式等进行优化,在实现数据调研与分析中,可从数据需求设计、ETL、数据建模、元数据管理等方面进行优化,在突破数据仓库与大数据融合的视角下,可为后续的数据应用于数据分析处理等提供参考依据。数据仓库与大数据的融合是在系统平台优化的基础上,对数据管理

过程、数据处理过程等方面进行优化,从而提高数据分析与业务处理的综合水平。

3.3 数据分析与数据整合

在创建 BI 层、数据管理以及源数据层的基础上,对不同数据关系进行整合与处理,可实现数据仓库技术的实际应用效果提升。在实际应用与处理的过程中,则需要对数据分析与业务处理的基础上,对数据关系进行分析,在数据建模与处理的前提下,利用数据仓库技术,对数据关系以及数据整合等方面进行调整,从而满足数据仓库建设与应用的综合发展需求。在数据仓库的搭建与分析中,可利用数据传输机制,对数据关系、信息处理过程等方式,落实数据访问服务。结构化、半结构以及非结构化数据的处理分析,则可以在 ETL 处理以及 Hadoop 平台,对不同数据关系进行调整与处理,在 BI 层实现数据分析与数据优化。在数据信息处理与分析的过程中,则需要利用数据挖掘的方式,对数据关系进行处理,从而提高数据分析水平。

4 结语

数据仓库与大数据的融合,则是在数据信息处理的基础上,对数据库技术的实际应用进行分析,在优化数据管理机制的前提下,可实现业务数据管理水平的进一步提升。数据仓库与大数据技术的融合,则是在数据挖掘与信息处

理的基础上,对数据管理过程、业务处理方式等方面进行优化,从而实现数据管理与数据信息处理效果提升。在大数据技术分析与业务处理的基础上,可通过技术融合的方式,对技术方案以及数据仓库等方面进行优化,通过数据分析处理与数据信息处理,建立数据架构,并建立数据仓库与大数据技术融合方案,从而实现大数据业务数据处理效果提升。

参考文献:

- [1] 刘强. 试论数据仓库与大数据融合 [J]. 电脑知识与技术, 2020, 16(10): 7-9.
- [2] 杨磊. 大数据的发展及数据仓库的融合应用 [J]. 数字技术与应用, 2019, 37(06): 62, 64.
- [3] 李岩岩, 潘文彬, 汪新. 数据融合技术在仓库监测系统中的应用 [J]. 蚌埠学院学报, 2016, 05(06): 5-8.
- [4] 于鹃. 数据仓库与大数据融合的探讨 [J]. 电信科学, 2015, 31(03): 166-170.
- [5] 万抒, 冯中华, 余文杰, 裴华. 针对攻击链的安全大数据多维融合分析架构和机制研究 [J]. 通信技术, 2021, 54(08): 1975-1980.
- [6] 周静, 余浩然, 米清奎, 黄忠, 龚伟. 统计数据与大数据源的融合应用 [J]. 重庆师范大学学报(自然科学版), 2021, 38(02): 77-83.

(上接第6页)

性, 移动互联网主题的移动模型和动态网络的高性能路由协议包括主机和子网多管阵列, 它位于相对稳定的位置, 被定义为特定前缀, 这是一个在理论上和算法上都非常优秀的移动对象, 可以为移动互联网建立动态中介网络。拓扑关系的频繁变化对分散网络以及路由实施有着非常大的影响, 消耗了大量的服务处理器与资源。除了保证动态网络路由协议之外, 还要解决动态拓扑管理存在的问题, 需要复杂的移动主机配置的示意图和移动操作的示意图, 便于找到子网内主机之间出现的子网传输与路由问题。为了保证路由的有效性, 需要设计路由协议, 分析动态代理网络所具有的新特征, 便于人们发现问题并总结路由协议在新环境下的理论以及技术上存在的缺陷, 实时了解动态, 便于路由协议根据实际情况及时修改协议设计和对象移动。

4. 研究移动互联网环境中的组播原理、算法和协议。组播理论、算法和技术不仅可以应用于现有的 IP 网络, 还可以应用于未来动态变化的网络, 为了积极应对网络发展情况, 需要研究动态路由协议理论。例如, 对现阶段的固定网络组播技术以及移动 IP 网络组播技术进行定义, 可以相应地管理和维护移动节点资源的网络变量。另外, 根据网络组播之间的相互协作, 为系统提供可靠的连接。在分析优化分层机制基础上, 解决组播与组播树以及地址管理等问题, 特别是节点切换、高速传输机制和组播路径问题的方法。组播路由协议是保证源组播子网到目标组播子网

的传送, 组播路由以树形为主, 分为最短路径树和共享树两种。本研究主要就是通过分层移动性管理来将节点与基本多播传输树分离, 避免移动节点对组播结构的影响, 并在变化网络中实现组播管理机制。

4 结语

当前, 计算机与移动互联网成为时代发展重要内容, 分析计算机与互联网技术的发展, 对企业发展有着重要作用。因此, 企业必须从技术发展的角度、意义和构想出发, 有效地整合当前发展过程的现状, 完善计算机和移动互联网技术水平, 满足时代发展需求。

参考文献:

- [1] 李响. 移动互联网背景下计算机技术的发展趋势探讨 [J]. 电脑知识与技术, 2018, 14(35): 39-40.
- [2] 魏景林. 计算机与移动互联网技术的发展 [J]. 电子技术与软件工程, 2017(12): 14.
- [3] 滑涛. 浅论计算机与移动互联网技术的发展 [J]. 科技经济导刊, 2016(14): 22.
- [4] 吴功宜, 吴英. 互联网、移动互联网和物联网技术发展与网络课程改革——对计算机专业研究生系统能力培养的思考与实践 [J]. 计算机教育, 2015(21): 107-110.