

电动汽车领域的大数据处理与分析

访国网北京市电力公司高级工程师潘鸣宇

■ 胡 颖/本刊记者



潘鸣宇 高级工程师
国网北京市电力公司

大数据的开发应用以及互联网思维的充分运用，将会推动电动汽车产业更加快速发展，创造美好的环境以及美好的生活。

随着电动汽车产业的不断发展以及相关信息系统的不断建设，电动汽车运行数据和运营数据不断增加，如何有效地集中存储分析这些数据，形成数据集中管理、资源统一共享的应用模式，并从中挖掘数据的潜在价值，是目前市场所需。为了对电动汽车领域的大数据处理与分析有更多的了解，本刊记者特约采访了有多年相关工作经验的国网北京市电力公司高级工程师潘鸣宇先生，请他就此方面问题给予解答。

电动汽车领域大数据的产生

电气应用：您好！非常感谢您百忙之中接受本刊的采访！电动汽车产业已进入大数据移动互联网时代，大数据的开发应用将会推动电动汽车产业的快速发展。首先，请您介绍一下电动汽车领域大数据是如何产生的？

潘鸣宇：电动汽车领域的的数据大致可分为两类：一类是电动汽车及相关设备的运行和设备检测数据；另一类是电动汽车产业运营数据。根据数据的内在结构，这些数据可以进一步细分为结构化数据和非结构化数据。结构化数据主要包括存储在关系数据库中的数据，目前电动汽车相关信息系统中的大部分数据是这种形式，随着电动汽车产业的不断发展，这部分数据增长很快。相对于结构化数据而言，不方便使用数据库二维逻辑表来表现的数据即称为非结构化数据，主要包括视频监控、图形图像处理等产生的数据。在电动汽车相关信息系统中，非结构化数据也占到了很大的比重，主要体现在电动汽车充电站监控系统中。

结构化数据根据处理时限要求又可以划分为实时数据和准实时数据，比如电动汽车充电桩远程控制、充电预约和无卡充电等数据是实时数据，需要快速、准确地处理；而大量的状态监测数据对实时性要求相对较

低，可以作为准实时数据处理。

电动汽车领域大数据的特点和关键技术

电气应用：根据您多年工作的积累，请您总结一下电动汽车领域大数据具有哪些特点？

潘鸣宇：电动汽车领域的大数据具备如下特征：

1) 数据规模巨大。仅以北京市为例，每天有上千辆电动汽车需要充电，近3 000台充电桩进行实时状态采集，日增长数据量达到太字节级。

2) 数据类型繁多。电动汽车领域数据分布广、种类众多，包括实时数据、历史数据、文本数据、多媒体数据和时间序列数据等各类结构化、半结构化数据以及非结构化数据，各类数据查询与处理的频度和性能要求也不尽相同。

3) 价值密度低。以视频为例，连续不间断监控过程中，可能有用的数据仅仅有1~2 s。在电动汽车充电站监控中存在同样问题，所采集的绝大部分数据都是正常数据，只有极少量的异常数据，而异常数据是状态检修的最重要依据。

4) 处理速度快。在几分之一秒内对大量数据进行分析，以支持决策制定。对在线状态数据的处理性能要求远高于离线数据。这种在线的流数据分析与挖掘同传统数据挖掘技术有本质的不同。

可以考虑为各类电动汽车相关数据引入一个新的属性，即数据的真实性。数据的真实性是指与特定类型数据相关的可靠性级别。高质量数据对于数据分析结果的正确性有重要影响，可通过数据筛选提高数据的真实性。

电气应用：大数据技术在电动汽车数据领域的运用将势在必行。请您详细介绍一下电动汽车领域有哪些大数据关键技术？

潘鸣宇：电动汽车领域包含如下几种大数据关键技术：

(1) 数据迁移技术

数据迁移技术可以实现从传统的数据库向分布式的非结构化数据库传递数据，并能够保证数据传输的可靠性、容错性和实时性，且提高数据传输的效率。在数据传输过程中，可以对数据进行简单的处理、过滤，提取有价值、可用数据，减少数据存储所需资源，提高后期数据分析的准确性。

(2) 数据分布存储技术

数据分布存储技术是基于分布式文件系统的非结构化数据存储，它可实现高并发的数据处理，确保数据存储的高效率、高安全性和高可靠性，并能最大程度地均衡利用所有系统资源。数据分布存储可采用下列方式实现：

1) 分布式文件系统。将数据存储于分布式文件系统中，可以保证数据的安全性，并能够在大数据量的情况下采用分布式处理思想，多节点同时处理数据。

2) 非结构化数据库。基于分布式文件系统的非结构化数据库，可以保

www.eage.com.cn

正泰 昆仑 系列
万能式断路器

实时温度传感器
持续提示运行状态
洞悉风险更可靠



CHINT



领先 每一步 可靠 新 高度 欢迎关注正泰微信公众号 [广告]

浙江正泰电器股份有限公司

股票代码：601877 400-817-7777 www.chint.net/kunlun

证高可靠性、高性能、面向列和可伸缩的特性,提供数据分析挖掘所必须的数据安全、高效率数据访问。

(3) 数据挖掘技术

数据挖掘技术实现针对数据的可靠性分析并进行挖掘处理,获取有价值的信息、事物间的关联和隐藏的信息等。数据挖掘技术可以分为以下几个过程:

1) 历史数据分析。采用已经在各案例中被成功使用并得到广泛认可的工具,针对大量数据进行处理,获取其中隐藏的信息,得到可靠的预期结果,并进行后期的数据处理。

2) 数据预测。根据已经获取的可用规律,通过当前所获取的数据,对未来进行可靠性的预测,实现数据挖掘分析的价值。

(4) 数据可视化技术

面对海量的数据,其中包含高准确度和高分辨度数据、时变数据以及多变量数据等,如何在有限的屏幕空间下,以一种直观、容易理解的方式展现给用户,是一项非常有挑战性的工作。数据可视化技术运用计算机图形学和图像处理技术,将数据绘制成高准确度、高分辨率的图像,并提供交互工具,有效利用人的视觉系统,允许实现改变数据处理和算法参数,对数据进行观察及定性和定量分析。通过此操作,可以直观形象地展示抽象的数据,在实现技术价值的同时以艺术的效果展示。

电动汽车领域大数据的处理和分析

电气应用: 可以通过哪些环节对电动汽车运行数据和运营数据进行分析处理? 各环节能够实现哪些功能?

潘鸣宇: 对电动汽车运行数据和运营数据的分析处理方式进行研究,改善现有数据分析处理方式,便于对充电集中时间、充电电量增量情况和充电电度变化情况等信息的统计、后续充电设施建设指导以及数据挖掘分析等提供一种基于电动汽车运行数据和运营数据的分析处理方法,其技术架构如图1所示。

本文利用目前流行的Hadoop框架体系,其核心技术包括分布式文件系统(HDFS)、分布式数据库(Hbase)、数据仓库工具(Hive)和处理过程(MapReduce)等。电动汽车充电运行数据和运营数据分析处理分为数据存储、数据整合、数据处理和数据分析四个环节。

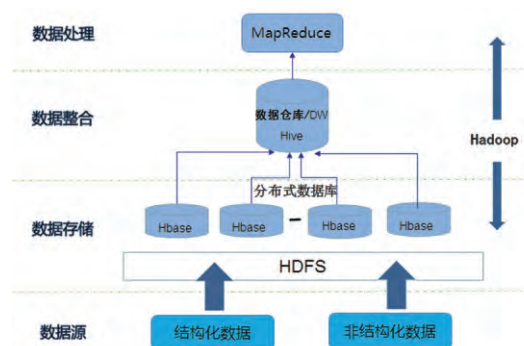


图1 技术架构

(1) 数据存储

电动汽车运行数据和运营数据包括结构化数据和非结构化数据。结构化数据包括充电桩状态代码、充电电压、充电电流、充电负荷、充电电量、充电时间和充电金额等;非结构化数据包括充电日志、充电视频等文件。Hadoop中的Hbase是一个分布式、面向列的开源数据库,它不同于一般的关系数据库,是一个适合于非结构化数据存储的数据库。因此可将电动汽车的非结构化数据存储于Hbase数据库中,由于该数据库是基于列存储的,有利于对数据库进行高效压缩,从而减少数据规模,因此有利于电动汽车的数据存储。

(2) 数据整合

由于电动汽车运行数据和运营数据来源比较分散且数据类型不统一,这就需要对数据进行整合。电动汽车数据整合技术利用基于Hadoop的数据仓库工具Hive搭建数据仓库,将Hbase数据库的数据全部存储至该数据仓库。该数据仓库存储方式是将结构化的数据文件映射为一张数据库表,并提供类SQL语言,实现完整的SQL查询功能。可以将SQL语句转换为MapReduce任务运行,十分适合数据仓库的统计分析。

(3) 数据处理

由于充电桩时钟跳变、异常充电和窃电等情况,造成电动汽车运行数据和运营数据存在异常,在对电动汽车数据进行挖掘分析前,需先进行数据处理,确保数据的准确性。

使用Hadoop的 MapReduce并行数据处理框架,MapReduce是一种简化的分布式编程模式,将复杂分布式处理过程分解为Map和Reduce过程,提高并发处理能力。让程序自动分布到一个由普通机器组成的超大集群上并发执行。它足够快速的批处理分析可以满足业务需求。

正泰昆仑系列

电动机控制与保护

70%-120%

吸合电压 无惧电压起伏 稳定运行更可靠

MapReduce致力于解决大规模数据处理的问题，因此在设计之初就考虑了数据的局部性原理，利用局部性原理将整个问题分而治之。MapReduce集群由普通PC构成，为无共享式架构。在处理之前，将数据集分布至各个节点，如图2所示。处理时，每个节点就近读取本地存储的数据处理，将处理后的数据进行合并、排序后再分发至Reduce节点，避免了大量数据的传输，提高了处理效率。无共享式架构的另一个好处是配合复制策略，集群可以具有良好的容错性，一部分节点的Down机对集群的正常工作不会造成影响。它是一个使用简易的软件框架，基于它写出来的应用程序能够运行在由上千个商用机器组成的大型集群上，并以一种可靠容错的方式并行处理上太字节级别的数据集。

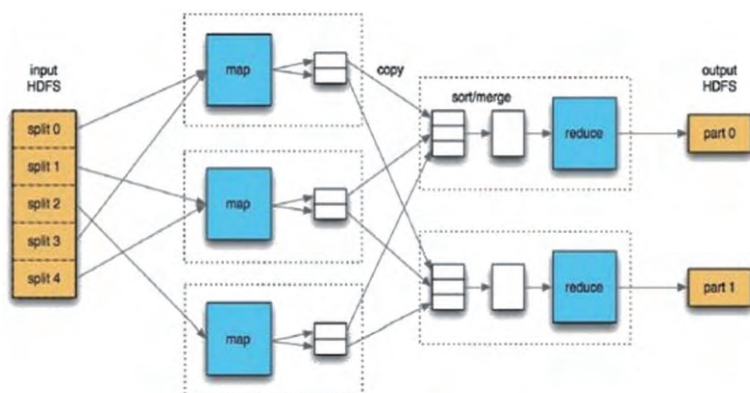


图2 无共享式架构

(4) 数据分析

目前，北京市电动汽车智能充换电服务网络运营管理系统已积累了大量的电动汽车用户运行数据和运营数据，通过运用本文中提到的大数据分析处理技术，实现对电动汽车充电站使用率、充电站月均负荷和充电站充电时间等数据的分析，持续跟踪观察电动汽车用户充电特征状况，探索电动汽车用户充电行为习惯，研究电动汽车用户充电消费模式，从而更好地开展电动汽车充电业务，提高用户服务的效率和质量。

电气应用：最后，请您谈谈电动汽车领域大数据技术未来的发展趋势。

潘鸣宇：随着电动汽车产业的不断发展、充电设施的不断建设以及电动汽车业务的不断推广，电动汽车的相关数据量将呈现迅速上升的态势。一般的数据分析已无法满足海量电动汽车数据上的应用，如何对多源异构的电动汽车数据进行挖掘分析，寻找其中的潜在价值引人深思。大数据技术在电动汽车数据领域的运用将势在必行。

电动汽车产业进入大数据移动互联网时代，应该用大数据的思维观念来处理数据，挖掘数据的潜在价值。电动汽车作为通信、计算机、电力电子、动力控制和驱动技术以及新材料技术等交通运输领域集成应用的产物，也为云计算、大数据和智能终端等新技术提供了率先应用的环境。大数据的开发应用以及互联网思维的充分运用，将会推动电动汽车产业更加快速发展，创造美好的环境以及美好的生活。EA

www.eage.com.cn



CHINT



领先每一步 可靠新高度 欢迎关注正泰微信公众号 [广告]

浙江正泰电器股份有限公司

股票代码：601877 400-817-7777 www.chint.net/kunlun