文章编号：2096-1472(2018)-09-05-02
DOI:10.19644/j.cnki.issn2096-1472.2018.09.002

基于智慧水务的供水大数据采集架构分析研究

朱炯名
(天津工业大学计算机科学与软件学院，天津 300387)

摘　要：本文主要论述的是“智慧水务”领域下供水大数据采集分析架构的设计，目前在供水数据采集方面，存在数据采集实时性低、海量数据存储困难和大数据计算复杂等难题。本文所设计的大数据采集分析架构，采用SOA架构理念，以面向服务的模式应用不同的互联网技术来解决以上难题，从而实现供水数据实时采集分析的业务需求。

关键词：智慧水务；SOA架构；大数据计算；实时采集分析
中图分类号：TP319　**文献标识码**：A

Research on Water Supply Big Data Collection Framework Based on *Intelligent Water*

ZHU Jiongming
(*School of Computer Science and Software Engineering*,*Tianjin Polytechnic University*,*Tianjin* 300387,*China*)

Abstract: This paper mainly discusses the design of the big data collection and analysis framework for water supply in the field of *Intelligent Water*.At present,there are problems in the data collection of water supply,such as low real-time data collection,difficult storage of large amounts of data and complicated calculation of big data.The big data collection and analysis architecture designed in this paper adopts the Service-Oriented Architecture (SOA) concept and applies different Internet technologies in a service-oriented mode to solve the above-mentioned problems,thus meeting the business needs of real-time collection and analysis of water supply data.

Keywords:*Intelligent Water*;SOA architecture;big data computing;real-time collection and analysis

1 引言(Introduction)

“智慧城市”的概念是当前高科技城市化建设的标志代名词，其实质便是使用先进的互联网信息化技术，通过各项关键信息的集中管控来实现城市智慧式的管理和运行，为城市中的人们创造更加便捷的生活设施和更美好的生活环境，从而促进城市的和谐可持续发展。“智慧城市”的发展所依赖的不仅仅是硬件设施的增强和完善，信息化建设更是重中之重，通过信息化建设串联城市中正常运行的各项基本要素，充分整合城市可使用的资源，通过物联网、大数据、云计算、人工智能等高科技信息技术手段来实现最优化配置，最精准化计算，以及最可靠的预测[1]。“智慧水务”是“智慧城市”中城市供水的关键一环，水资源的供求关系对于一个城市来讲至关重要，如何有效的避免水资源的过度浪费，精准管控水源地的水质水况，都是“智慧水务”的关键性研究目标，通过对各地区供水数据研究发现，供水情况复杂多变，各地区之间水源地、地质地貌各有不同，那就需要对不同城市的供水循环数据进行有效的分析而制定出一套具有全面的水务数据分析架构，再进行个性化定制，从而达到因地制宜的效果[2]。大数据分布式计算框架便完全可胜任水务数据的分析计算。

2 相关工作及背景知识(Related work and background information)

2.1 Spark框架

Spark是由UC Berkeley AMP Lab(加州大学伯克利分校的AMP实验室)所开源出来的大数据分布式计算框架，现已发展为Apache的孵化项目。Spark出色的计算效率被越来越多的开发者所推崇，在内存计算下要比Hadoop快100倍，并且具有很高的易用性和通用性。Spark具有很强的适应性，能够读取HBase、HDFS、Mysql、MongoDB等多种类型数据库的数据，Spark借鉴了MapReduce的优点同时改进了MapReduce的明显缺陷，是目前最为流行的大数据计算框架之一。并且Spark用着自身一套庞大的生态系统[3]，图1为Spark生态系统图。

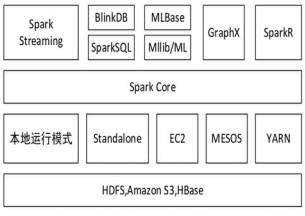

图1 Spark生态系统
Fig.1 Spark ecosystem

2.2 MongoDB文档数据库

在面对Web 2.0兴起的浪潮中，高并发、大数据量的应用使得关系型数据库变得力不从心，Nosql数据库此时兴起，Nosql代指非关系型数据库，抛开了传统关系型数据库的设计，从更多角度去解决高并发和海量数据所带来的压力。MongoDB则是Nosql中的佼佼者，是一种基于文档设计的非关系型数据库，可用于海量数据存储。MongoDB的数据存储格式使用类似JSON数据格式的变种BSON，其格式如：{“data”：“test”，“value”：“123”}。MongoDB最大的特点在于其所支持的查询语言非常强大，提供多种组合式查询，并且提供了多种语言的底层API。同时MongoDB支持集群化部署，在资源有限的情况下MongoDB还支持分片技术来应对高吞吐量、单机扩容困难的情况，如图2所示，一个拥有10TB的数据库，在使用分片技术后，划分为五个2TB的分片，当查询某一条数据时，仅仅访问其所在的分片即可，有效节省了内存消耗[4]。

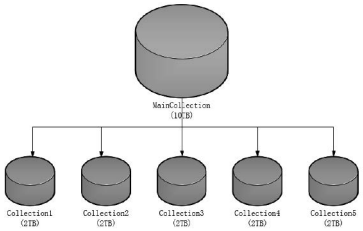

图2 MongoDB分片技术

Fig.2 MongoDB sharding technology

2.3 Netty网络通信框架

Netty是一个支持异步的，事件驱动的网络应用程序框架，也就是说Netty是一个基于NIO的客户端和服务端的通讯开发框架，支持多种网络通讯协议。抛开Socket繁杂的编程体验，Netty提供更加轻便的编程API，Netty针对多种传输类型设计了统一的接口，即阻塞和非阻塞的实现，同时内部实现了更简单更强大的线程模型，实现了真正的无连接的数据报套接字的支持[5]。Netty框架对TCP请求支持可达到百万级，请求响应效率更高，使用Netty作为GPRS底层采集设备的服务端是一个非常好的选择。图3为Netty架构图。

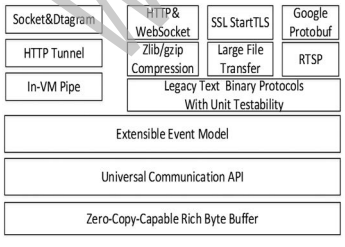

图3 Netty框架结构

Fig.3 Netty framework

2.4 Redis缓存数据库

上文中提到过Nosql数据库的相关概念，Redis数据库同样是Nosql数据库中的一员，与MongoDB不同的是，Redis是一个基于内存的数据库，采用Key-Value键值对的形式保存数据，由于其基于内存的特性，Redis对于数据的存储和读取具有非常高的效率。同样的，Redis同时支持集群化部署，保证了集群环境下的数据统一。Redis基于内存的特性也使得其成为目前非常流程的缓存数据库，搭配缓存机制来使用Redis不仅能够达到数据缓存的效果，同时根据不同的业务场景，能够将数据进行持久化操作，达到了缓存和持久化统一结合的效果[6]。

3 数据采集(Data collection)

“智慧水务”的核心在于供水数据的实时采集和数据的智能化分析预测，如何能够达到资源的最优化分配取决于供水数据的分析结果，如何实时准确地获取供水数据则是“智慧水务”大数据采集架构的核心研究问题[7]。首先实时采集数据存在两个关键点：一是时效性，二则是海量数据存储性能，能够顺利解决二者，那么其他问题均可迎刃而解。根据上述技术架构，使用MongoDB作为用来存储水务数据，其性能和分片扩展能够完美应对实时的水表采集数据；Spark大数据计算框架用来实时计算采集数据，提供完整有效的计算结果，并且根据历史分析记录，能够精准判断当前是否存在报警或爆管等异常现象，达到预测效果；Netty用作通讯服务，将远控解析协议进行解析拼装后下发至采集器或智能水表，返回数据后完成一次解析后存储MongoDB中。

4 技术架构(Technology architecture)

“智慧水务”大数据采集系统架构是一套完整的囊括了大数据计算体系，实时通讯体系的互联网架构。首先上层使用Nginx做负载均衡，用户访问通道统一，使用Nginx反向代理减轻服务端压力，同时使用Redis作为缓存库，对访问频繁的历史信息进行有效缓存和更新，提供给用户更好的操作体验和更短的响应时长。

Mysql用于营收存储关系型数据，例如用户基础信息、水表基础信息、工作人员基础信息等，MongoDB中用于存储智能水表的采集数据，由于智能水表的采集数据量大，实时性高，使用Spark大数据分析框架对采集数据进行实时计算，计算后的数据存储到结果MongoDB中[8]，集中管控云平台可实时调用MongDB查询相关采集统计数据[6]。

Netty通信服务接收管理平台的采集指令，为避免信道阻塞，这里同样适用Nginx做负载，同时调用MQ消息队列对水表进行指令队列式推送，从而保证采集命令能够有效下发。水表协议转换交由Netty端进行输入输出转换，水表响应指令返回数据后，由通信服务转换后直接存储到MongoDB中。根据Netty架构模型TCP通讯会建立相应的Channel，为保证Channel的时效性和定期清理规则，Redis是用于缓存Channel，Redis可以自定义缓存时效，也可定义是否需要持久化，建立连接的Channel缓存至Redis中，当消息接收或发送结束，Netty为节省资源，将制定Channel销毁，同时删除Redis中的缓存，达到了预期的缓存效果[9]。图4为整体采集架构方案图。

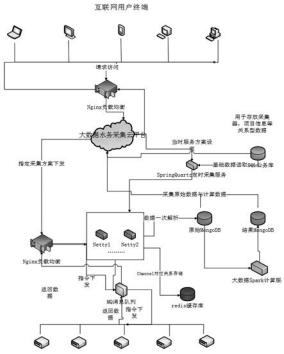

图4 采集框架结构设计

Fig.4 Collection framework design

5 结论(Conclusion)

本文通过对“智慧水务”的供水数据采集和数据分析为研究基础，设计提出了一套完整的水务大数据采集分析框架，一方面解决数据采集的实时性问题，另一方面解决大数据的存储和分析问题。整体框架中不仅仅使用到了Spark、MongoDB、Redis、Netty这些成熟先进的开发框架和数据库，同时也引入了SOA面向服务的设计模式，将各应用模块进行解耦剥离，各模块利用其优势专注于需求服务，各服务之间设计相应的接口调用，最终实现了一套完整的“智慧水务”技术架构体系。

参考文献(References)

[1] 张义,陈虞君,杜博文,等.智慧城市多模式数据融合模型[J].北京航空航天大学学报,2016,42(12):2683-2690.

[2] 杨明祥,蒋云钟,田雨,等.智慧水务建设需求探析[J].清华大学学报(自然科学版),2014,54(01):133-136;144.

[3] Holden Karau.余璜,张磊,译.Spark快速数据处理[M].北京:机械工业出版社,2014:47-84.

[4] Kristina Chodorow,Michael Dirolf.程显峰,译.MongoDB权威指南[M].北京:人民邮电出版社,2011:23-87.

[5] 陈强,武佳佳.基于Netty高并发物联网服务器研究与设计[J].电子技术与软件工程,2018(07):34-35.

[6] Haber,Itamar.MongoDB and Redis pair volume with velocity[EB/OL].https://www.infoworld.com/article/3008052/nosql/mongodb-and-redis-pair-volume-with-velocity.html,2015-12-01.

[7] Gu Jifa.Data,Information,Knowledge,Wisdom and Meta-Synthesis of Wisdom-Comment on Wisdom Global and Wisdom Cities[J].Procedia Computer Science,2013(17):713-719.

[8] 章毅,郭泉,王建勇.大数据分析的神经网络方法[J].工程科学与技术,2017,49(01):9-18.

[9] 马豫星.Redis数据库特性分析[J].物联网技术,2015,5(03):105-106.

作者简介：

朱炯名(1990-)，男，硕士，工程师.研究领域：大数据计算，人工智能.

(上接第17页)

4.2 界面实现

移动学习小程序的主界面如图7所示。

图7 移动学习小程序主界面

Fig.7 Main interface of mobile learning applet

5 结论(Conclusion)

目前该小程序已设计完成，正在试用期间，共上线三门课程，342名学生注册使用。在试用过程中，系统使用流畅。基于微信小程序的移动学习平台实现了预期的功能，但也发现了一些有待完善的地方，比如，在答题游戏的设计上，一些相关参数的设置(比如：闯关级数、题目数，题目的难易程度等)仅凭借教师的经验处理，还缺少数据的支持，出现有学生反馈游戏环节吸引力不够的问题。未来在采集到比较充分的数据后，还需要结合统计学分析做进一步改进。另外，微信小程序要求程序尽量简洁，程序使用流畅，未来在代码设计上还需要进一步优化。

参考文献(References)

[1] 苏丹蕊,李玉斌.近五年我国移动学习研究综述[J].中国教育技术装备,2016(22):8-10.

[2] 宋巍.H.264技术在移动学习平台中的应用[J].中国远程教育,2014(2):87-90.

[3] 刘萍,丁萍莉.基于微信小程序的互联网+教育学习平台的研究与设计[J].内江科技,2016(12):45-46.

[4] 朱玉强.微信小程序在图书馆移动服务中的应用实践_以排架游戏为例[J].图书馆论坛,2017,37(7):132-138.

[5] 李显明,牛琳,马燕,等.基于微信小程序的医嘱服务研究[J].计算机时代, 2018(1):66-68;73.

作者简介：

牛　琳(1978-)，女，硕士，讲师.研究领域：信息管理系统，数据挖掘.

吴　琳(1978-)，女，硕士，讲师.研究领域：微生物.