

基于微服务架构的大数据应用开发创新实践

杨 强, 张钧鸣

(江苏核电有限公司, 江苏 连云港 222042)

摘 要:为解决核电企业面临的海量数据组合查询搜索问题,江苏核电积极开展基于大数据技术和面向微服务架构的应用架构研究。以开源软件体系为依托,采用分层架构设计的方法,将大数据技术平台划分为数据采集层、数据计算层、数据服务层、数据应用层这几个层次。在各个层次上,研究和探索使用主流的大数据技术产品满足对应层次的需要。在应用架构上,创新性使用微服务架构来替代以往的单体应用开发模式,基于业内主流的 Spring Cloud 技术体系,实现了应用架构的全面微服务化。通过新的应用架构,不仅通过用户实时统计分析的应用开发来验证了微服务相关的服务注册中心、服务配置中心、服务发现、网关路由、服务高可用、服务负载均衡、服务快速 API 化等相关技术方案,而且为基于大数据与微服务的技术融合提供了一条可工程化的路径。

关键词:大数据;微服务;数字核电;架构设计;应用创新

文章编号:2096-4633(2019)03-0071-06 **中图分类号:**TM623 **文献标志码:**B

1 背景

大数据、人工智能、物联网等新一代信息技术已经成为各行各业乃至国家层面的重要创新依托。两化融合和智能制造等概念的热潮给传统企业的信息化发展注入了新的动力,也来了新的挑战。企业信息化人员的职责不仅需要对当前运作满足内部业务流程管理需要的信息系统负责,还需要思考与探索新兴信息技术在传统企业的应用方式和方法。如何快速地将企业内部的业务需求通过新一代信息技术实现落地并能跟随用户的使用进行持续性的优化完善成为企业信息化工作人员关注的中心。因此,核电企业的信息化工作人员一直在探索一种能够在资源有限的情况下支撑起快速发展的业务,解决大数据环境下信息系统的建设问题^[1]。

目前电力行业大数据已经广泛应用于配电网和电力设备在线监测中,如配电网运行状态诊断、用户用电行为特征分析、输电线路状态评估、基于复杂相关关系识别的输变电设备故障诊断和预测技术^[2-3]、基于大数据和信息融合的输变电设备关键性能动态评估及预测等。虽然电力行业已经产生了海量数据,但大多数仅仅实现了数据存储、查询、统计等最基层的功能,尚未深入挖掘出隐藏在海量数据背后潜在的价值。若需要发现电力大数据中蕴含的更深层次的规律,保证系统运行的经济性、安

全性和可靠性,需要通过数据挖掘技术来完成决策任务。大数据是业务驱动型技术,以巨量数据为支撑,根据业务需求,挖掘数据关联性,为业务提供决策支持^[4]。

2 相关概念介绍

双模 IT 概念由 Gartner 提出,双模 IT 是指两种不同的 IT 工作模式。一种模式专注于可预测性,其目标是可靠、可用和低成本,适用于需求明确的工作。比如 ERP 实施项目,在需求分析阶段就必须将用户的需求明确下来。另一种模式是探索性的工作,需求在开始阶段并不明确。比如用大数据分析设备缺陷原因或者利用机器学习来预测设备劣化趋势,在项目开始阶段,用户只知道项目的目标,对于具体的需求并不明确,需要和 IT 部门一起来探索。对于核电企业这种以资产运营为核心的企业,必须通过将双模 IT 实施,才能在数字化浪潮中打赢数字化转型这场硬仗^[5-6]。

大数据指的是所涉及的资料量规模巨大到无法通过目前主流软件工具,在合理时间内达到摄取、管理、处理、并整理成为帮助企业经营决策更准确的资讯。大数据有数据量大、数据种类多、要求实时性强、数据所蕴藏的价值大等特点,简称为 4V,即 Volume、Velocity、Variety、Veracity。大数据的数据来源可能是多种来源,这些来源包括搭载感测设备的

移动设备、高空感测科技(遥感)、软件记录、相机、麦克风、无线射频辨识(RFID)和无线感测网络。我们通常声称的大数据应用是指利用先进的大数据管理和处理技术,针对实际业务所面临的日益膨胀的海量数据存储问题以及在此之上的海量数据处理问题而设计相应的解决方案并应用^[7]。

微服务架构是相对于传统的整体式架构(Monolithic Architecture)而言的。传统的应用开发系统,如 CRM、ERP 等大型应用,随着新需求的不断增加,企业更新和修复大型整体式应用变得越来越困难,影响到企业应用功能的快速上线;随着应用云化的日益普及,需要采用传统 IT 不同开发运维模式来开展应用的开发过程。微服务架构是一种架构风格,一个大型复杂软件应用由一个或多个微服务组成。系统中的各个微服务可被独立部署,各个微服务之间是松耦合的。每个微服务仅关注于完成一件任务并很好地完成该任务。在所有情况下,每个任务代表着一个小的业务能力^[8]。

3 基于大数据与微服务的应用架构创新案例

3.1 业务场景介绍

核电厂现场检修是保持设备高可用的重要业务之一。目前的主要工作业务流程一般如图 1 所示。各个渠道的缺陷申请在提出后,会由公司级的计划管理人员分发到具体的业务处室,业务处室的计划人员通过评估,分派到具体的执行班组,执行班组的工作负责人准备工作开工的相关文件、申请工作需要备品备件和工器具和检查对应的可用性、协调需要申请的工作隔离许可、申请工作的执行窗口。在工作任务执行时,处室的管理人员需要了解工作负责人的工作申请执行情况,例如具体的工作负责人的工作负载、工作负责人的人员来源分布、工单耗时情况等等。

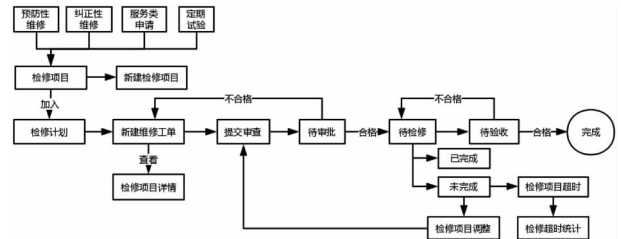


图 1 核电厂维修过程示意图

Fig. 1 Diagram of maintenance process in nuclear power plant

3.2 总体设计

在本应用案例中,依托公司已经建立的大数据平台,实现对数据的采集、计算和服务。大数据技术平台将作为企业所有信息资源的集中管理平台,实现精细化管理和业务创新。平台中的大部分功能组件采用主流的开源软件产品,无现成软件产品的部分通过公司自主开发来实现。总体设计如图 2 所示。

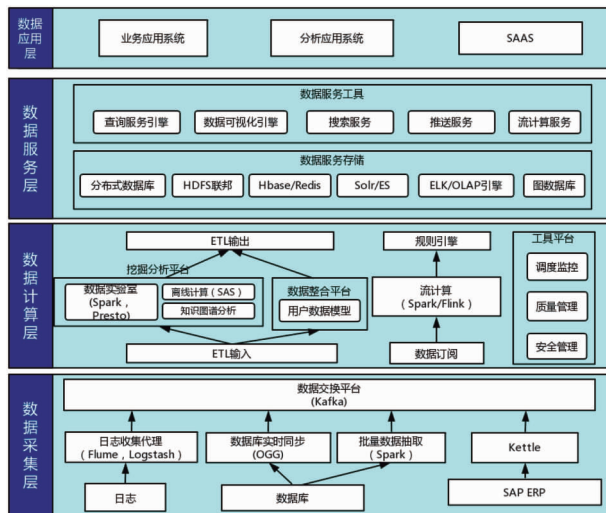


图 2 大数据平台的总体设计图

Fig. 2 Overall design of big data platform

总体设计分为数据采集层、数据计算层、数据服务层和数据应用层。^[9]数据采集层将数据库、SAP 和日志等数据源的数据加载到数据交换平台,不同的数据源采用不同的采集方案,例如日志文件采用 Flume 或者 Logstash 能够开箱即用的使用,且在实现效率有较高的性能。数据计算层将 ETL 的输入作为计算的来源,对于加载的数据直接开展计算过程,计算可以分为离线计算、实时计算,计算上可以开展自然语言处理和知识图谱分析等工作。在这个层面也包括对整体的处理流程的调度监控、大数据的质量管理和大数据相关的安全管理。数据服务层分为数据存储服务和数据服务工具,数据存储服务包括采用分布式数据库、分布式文件系统、内存数据库、搜索引擎、OLAP 引擎以及当先比较流行的图数据库。在数据应用层,根据面向的用户不同,主要分为业务应用系统、分析应用系统和 SAAS 为代表的云应用^[10]。

3.2 大数据关键技术与实现

在数据采集上,工作申请和维修工单存在于 SAP ERP 系统之中,因此这个部分需要解决的问题

是利用数据采集工具从源系统中抽取数据并输入到计算层中。本案例使用 Kettle 软件, Kettle 是开源的 ETL 工具, 产品基于 Java 可以运行在任何计算机上, Kettle 易于开发且避免了自定义开发, 所有功能都通过用户界面完成, 没有命名规则要求, 全过程对用户透明, 具有灵活的数据通道, 只映射需要映射的字段。Kettle 的内容包括转换和作业, 转换是 ETL 解决方案中最主要的部分, 它处理抽取、转换、加载各阶段各种对数据行的操作。转换包括一个或多个步骤, 如读取文件、过滤输出行、数据清洗或将数据加载到数据库。

在数据计算上, 由于从数据采集层过来的数据将会自动发送到 Kafka 序列, 因此我们需要一个跟 Kafka 能够较好集成的计算框架。Kafka 是由 Apache 软件基金会开发的一个开源流处理平台, 由 Scala 和 Java 编写, 目标是为处理实时数据提供一个统一、高吞吐、低延迟的平台。其持久化层本质上是一个“按照分布式事务日志架构的大规模发布/订阅消息队列”, 这使它作为企业级基础设施来处理流式数据非常有价值最终我们选择了 Spark Streaming。这个框架依托于 Spark, Spark 是一个类似于 MapReduce 的分布式计算框架, 其核心是弹性分布式数据集, 提供了比 MapReduce 更丰富的模型, 可以在快速在内存中对数据集进行多次迭代, 以支持复杂的数据挖掘算法和图形计算算法。Spark Streaming 是一种构建在 Spark 上的实时计算框架, 它扩展了 Spark 处理大规模流式数据的能力。Spark Streaming 的优势在于它能运行在 100 + 的结点上, 并达到秒级延迟; 使用基于内存的 Spark 作为执行引擎, 具有高效和容错的特性; 能集成 Spark 的批处理和交互查询; 为实现复杂的算法提供和批处理类似的简单接口^[11]。

在数据服务层中的数据存储上, 采用分布式数据库产品 HBase, HBase 是一个分布式的、面向列的开源数据库, 该技术来源于 Fay Chang 所撰写的 Google 论文“Bigtable: 一个结构化数据的分布式存储系统”。就像 Bigtable 利用了 Google 文件系统 (File System) 所提供的分布式数据存储一样, HBase 在 Hadoop 之上提供了类似于 Bigtable 的能力。HBase 是 Apache 的 Hadoop 项目的子项目。HBase 不同于一般的关系数据库, 它是一个适合于非结构化数据存储的数据库。另一个不同的是 HBase 基

于列的而不是基于行的模式。利用 Hadoop HDFS 作为其文件存储系统, 利用 Hadoop MapReduce 来处理 HBase 中的海量数据, 利用 Zookeeper 作为其分布式协同服务, 主要用来存储非结构化和半结构化的松散数据 (列存 NoSQL 数据库)^[12]。

在数据服务层的搜索服务上, 采用 Apache Solr 作为搜索服务器, 它支持 REST 风格 API。Solr 是基于 Lucene 的, Lucene 支持强大的匹配能力, 如短语, 通配符, 连接, 分组和更多不同的数据类型。Solr 使用 Apache Zookeeper 特别针对高流量进行优化。Apache Solr 提供各式各样的功能, Solr 提供了先进的全文搜索功能, 支持开放接口标准例如 XML, JSON 和 HTTP, 具有高度的可扩展和容错性。该服务支持模式和无模式配置以及分页搜索和过滤。Solr 的搜索服务支持通过首字母查询、支持通过通配符查询、按名称查询等多种搜索条件。^[13-15]

HBase 的一级索引是 rowkey, 只能通过 rowkey 进行检索。如果需要对 HBase 里面列族的列进行一些组合查询, 就需要采用 HBase 的二级索引方案来进行多条件的查询。基于 Solr 的 HBase 多条件查询原理很简单, 将 HBase 表中涉及条件过滤的字段和 rowkey 在 Solr 中建立索引, 通过 Solr 的多条件查询快速获得符合过滤条件的 rowkey 值, 拿到这些 rowkey 之后在 HBase 中通过指定 rowkey 进行查询^[16]。

3.3 微服务架构的实现

本案例的应用开发框架采用 Spring Boot, 这个框架的设计目的是用来简化新 Spring 应用的初始搭建以及开发过程。Spring Boot 框架使用了特定的方式来进行配置, 从而使开发人员不再需要定义样板化的配置。框架能够快速构建独立的 Spring Application, 使用内嵌容器, 无需部署到外部容器。框架提供了众多扩展的 ‘starter’, 通过依赖感知自动配置, 你要做的仅仅是添加 Maven 依赖。框架提供生产环境下的性能健康状态监控及外部化配置, 提供实时的监控。框架无需生成代码及 XML 配置。能够很快的投入使用。

在微服务架构的实现上, 主要使用 Spring Cloud。Spring Cloud 是一系列框架的有序集合。它利用 Spring Boot 的开发便利性巧妙地简化了分布式系统基础设施的开发, 如服务发现注册、配置中

心、消息总线、负载均衡、断路器、数据监控等,都可以用 Spring Boot 的开发风格做到一键启动和部署。Spring Cloud 并没有重复制造轮子,它只是将目前各家公司开发的比较成熟、经得起实际考验的服务框架组合起来,通过 Spring Boot 风格进行再封装屏蔽掉了复杂的配置和实现原理,最终给开发者留出了一套简单易懂、易部署和易维护的分布式系统开发工具包。^[17]

Eureka 是一个 RESTful 服务,使用 REST 进行通讯。Eureka 用来运行中间层服务。由两个组件组成。Eureka 服务器用作服务注册服务器。Eureka 客户端是一个 java 客户端,用来简化与服务器的交互、作为轮询负载均衡器,并提供服务的故障切换支持。

Feign 是一个声明式的 web 服务客户端;它支持可插拔的注解,包含 Feign 注解和 JAX-RS 注解;并且支持可插拔的编码、解码;Feign 集成了 Ribbon 和 Eureka 来提供对 http 客户端的负载均衡。

Zuul 是在云平台上提供动态路由,监控,弹性,安全等边缘服务的框架。相当于是设备和流应用的 Web 网站后端所有请求的前门。Zuul 实现了跨横切面的功能:如路由,安全认证,容错,限流,日志等。

本案例的架构图如图 3 所示,设计了统计 API 微服务、基础数据服务微服务和统计应用前端服务。每个服务都注册到 eureka 服务中心,且微服务使用 feign 作为客户端。通过自主设置服务的实例数目,可以轻松实现微服务的高可用和负载均衡^[18-19]。

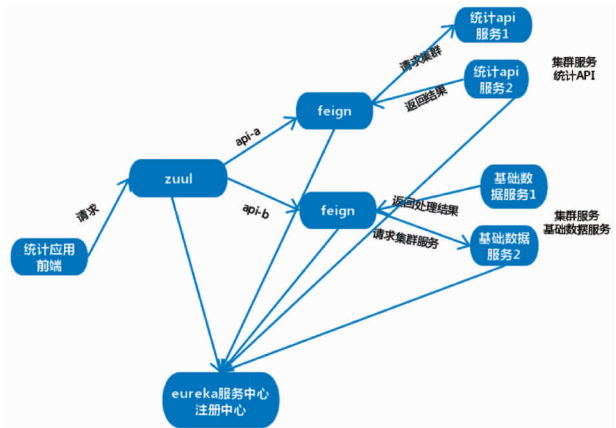


图 3 应用的微服务架构设计图

Fig. 3 Diagram of Microservice architecture design for application

3.4 应用效果

本案例投用后,用户可以在很短的时间里面从海量源数据中获取自己需要的统计数据,如图 4 所

示,用户可以根据工单的日期、机组、工作类型、设备等级、执行部门、工作班组、工作负责人等搜索项进行自由组合,查询出自主维修数量比例和自主维修工时比例,用户点击具体的数字可以查看到详细的工作申请的数据清单。38 万条检修数据的多维统计及明细加载只需 3 秒。



图 4 应用的实施效果图

Fig. 4 Diagram of Application Implementation

4 总结与展望

核电企业为解决实际遇到的海量数据组合查询搜索问题,积极开展基于大数据技术和面向微服务架构的应用架构研究。按照分层架构对大数据技术平台进行设计,将平台划分为数据采集层、数据计算层、数据服务层、数据应用层。在各个层次,使用主流的大数据技术解决对应层次的需要。在应用架构上,使用 Spring Cloud 实现应用的微服务化,实现了服务注册、服务发现、网关路由、服务高可用、服务负载均衡等。通过对相关架构的设计和新兴技术的实践,满足了用户实时统计分析的需求。

参考文献:

[1] 袁捷. 贵州电网大数据应用探讨[J]. 电力大数据,2017,20(12):4-7.
YUAN Jie. Discussion on application of big data in Guizhou power grid[J]. Power Systems and Big Data,2017,20(12):4-7.
[2] 张冰玉. 基于数据挖掘技术的短期电力负荷预测[J]. 电力大数据,2017,20(10):18-21.
ZHANG Bingyu. Prediction of short-term power load based on data mining technology[J]. Power Systems and Big Data, 2017, 20(10):18-21.
[3] 李占英. 智能配电网大数据应用技术与前景分析[J]. 电力大数据,2017,20(11):18-20.
LI Zhanying. Intelligent power distribution network and prospect analysis technology of data application[J]. Power Systems and Big Data,2017,20(11):18-20.
[4] 曾四鸣. 大数据挖掘技术在电力行业中的应用[J]. 电力大数

- 据,2017,20(09):81-84.
- ZENG Siming. The application of big data mining(DM)technology in power system[J]. Power Systems and Big Data,2017,20(09):81-84.
- [5] 张楠. Gartner 陈勇释疑双模 IT[N]. 中国计算机报,2016-10-17(006).
- [6] 王盈. 双模 IT:未来企业的必走之路[J]. 软件和集成电路,2016(11):12-13.
- WANG Ying. Dual-mode IT:the must-go path for future enterprises [J]. Software and Integrated Circuits,2016(11):12-13.
- [7] 柏秋云. 大数据的价值与挑战[J]. 科技信息,2013(17):479.
- Bai Qiuyun. The value and challenge of big data[J]. Science & Technology Information,2013(17):479.
- [8] 李贞昊. 微服务架构的发展与影响分析[J]. 信息系统工程,2017(01):154-155.
- LI Zhenhao. Development and impact analysis of microservice architecture[J]. Information Systems Engineering,2017(01):154-155.
- [9] 饶玮,蒋静,周爱华,等. 面向全球能源互联网的电力大数据基础体系架构和标准体系研究[J]. 电力信息与通信技术,2016,14(04):1-8.
- RAO Wei,JIANG Jing,ZHOU Aihua,et al. Research on power big data basic architecture and standard system for global energy interconnection[J]. Electric Power Information and Communication Technology,2016,14(04):1-8.
- [10] 朱朝阳,王继业,邓春宇. 电力大数据平台研究与设计[J]. 电力信息与通信技术,2015,13(06):1-7.
- ZHU Chaoyang, WANG Jiye, DENG Chunyu. Research and design of electric power big data platform [J]. Electric Power Information and Communication Technology, 2015, 13 (06): 1-7.
- [11] 岑凯伦,于红岩,杨腾霄. 大数据下基于 Spark 的电商实时推荐系统的设计与实现[J]. 现代计算机(专业版),2016(24):61-69.
- CEN Kailun, YU Hongyan, YANG Tengwei. Design and implement of e-commerce real-time recommender system with spark based on big data [J]. Modern Computer,2016(24):61-69.
- [12] 孟祥萍,周来,王晖,等. HBase 在智能电网异构数据同步中的应用[J]. 电力系统保护与控制,2015,43(24):122-128.
- Meng Xiangping, Zhou Lai, Wang Hui, et al. Applications of Hbase for heterogeneous data synchronization in smart grid [J]. Power System Protection and Control,2015,43(24):122-128.
- [13] 程知群,章超,韩高帅. 基于 Solr 的数据检索技术研究[J]. 杭州电子科技大学学报(自然科学版),2017,37(01):11-15.
- CHENG Zhiqun, ZHANG Chao, HAN Gaoshuai. Data retrieval technique research based on solr [J]. Journal of Hangzhou Dianzi University (Natural Science),2017,37(01):11-15.
- [14] 李戴维,李宁. 基于 Solr 的分布式全文检索系统的研究与实现[J]. 计算机与现代化,2012(11):171-176.
- LI Daiwei, LI Ning. Research and implementation of distributed full-text retrieval system based on solr [J]. Computer and Modernization,2012(11):171-176.
- [15] 霍庆,刘培植. 使用 Solr 为大数据库搭建搜索引擎[J]. 软件,2011,32(06):11-14.
- HUO Qing, LIU Peizhi. Use solr to build a search engine for large database [J]. Software,2011,32(06):11-14.
- [16] 王文贤,陈兴蜀,王海舟,等. 一种基于 Solr 的 HBase 海量数据二级索引方案[J]. 信息网络安全,2017(08):39-44.
- WANG Wenxian, CHEN Xingshu, WANG Haizhou, et al. A secondary index scheme of big data in HBase based on Solr [J]. Information Network Security,2017(08):39-44.
- [17] 王方旭. 基于 Spring Cloud 实现业务系统微服务化的设计与实现[J]. 电子技术与软件工程,2018(08):60-61.
- WANG Fangxu. Design and implementation of business system microservice based on spring cloud [J]. Electronic Technology and Software Engineering,2018(08):60-61.
- [18] 马雄. 基于微服务架构的系统设计与开发[D]. 南京:南京邮电大学,2017.
- [19] 洪华军,吴建波,冷文浩. 一种基于微服务架构的业务系统设计与实现[J]. 计算机与数字工程,2018,46(01):149-154.
- Hong Huajun, Wu Jianbo, Leng Wenhao. Business system design and implementation based on microservice architecture [J]. Computer and Digital Engineering,2018,46(01):149-154.

收稿日期:2018-12-06

作者简介:



杨强(1988),男,硕士,高级工程师,主要从事核电厂信息化方面的研究工作。

(本文责任编辑:龙海丽)

Innovation practice of big data application development based on micro service architecture

YANG Qiang,ZHANG Junming

(Jiangsu Nuclear Power Co. ,Ltd. ,Lianyungang 222042 Jiangsu,China)

Abstract:In order to solve the massive data query query problem faced by nuclear power enterprises,JNPC carried out an research on application architecture based on big data technology and microservice architecture. Based on the open source software system and adopting the layered architecture design method,the big data technology platform is divided into several levels:data acquisition layer,data calculation layer,data service layer and application layer. At all levels,JNPC research and explore the use of mainstream big data technology products to meet the needs of the corresponding level. In the application architecture,the microservice architecture is innovatively used to replace the previous single application development model. Based on the mainstream Spring Cloud technology system in the industry,the application architecture is fully microservice. Through the new application architecture,JNPC not only implemented technical solutions such as service registration center,service configuration center,service discovery,gateway routing,service high availability,service load balancing,service rapid APIlization,etc. but also provide an engineerable path for technology convergence based on big data and microservice.

Key words:big data;microservice;digital nuclear power;architecture design;application innovation