



运营技术广角

电信运营商物联网客户运营分析 系统的构建及应用

张勋, 翟春辉, 张呈宇, 贾捷, 龙岳

(中国联合网络通信有限公司研究院, 北京 100176)

摘要: 简要分析了中国联通在物联网客户运营能力上的挑战与诉求, 结合中国联通研究院的大数据分析能力, 提出运用能力构建完善标准体系, 推动物联网规模应用, 实现物联网客户运营大数据化的目标。基于电信行业物联网运营模式的实践经验, 描述了如何以大数据能力为基础进行面向运营商的更为精细化的企业客户运营, 从技术创新的角度实现客户运营服务的精准化、高效化。以提升客户数据运营的准确性和前瞻性为目的, 探索如何利用数据分析能力, 构建精准的客户运营支撑体系。最后, 讲述了大数据应用于客户运营场景的重要价值和对客户运营未来的展望。

关键词: 物联网; 客户运营; 大数据

中图分类号: TP319

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2019141

Construction and application of telecom operator IoT customer operation analysis system

ZHANG Xun, ZHAI Chunhui, ZHANG Chengyu, JIA Jie, LONG Yue

Research Institute of China United Network Communication Co., Ltd., Beijing 100176, China

Abstract: China Unicom's challenges and appeals in the operation capabilities of IoT customers were briefly analyzed. Combined with the big data analysis capabilities of China Unicom Research Institute, a complete standard system was built. The scale of IoT applications was promoted and the goal of IoT customer operations big data was realized. Based on the practical experience of the IoT operating model in the telecom industry, how to implement more refined enterprise customer operations for operators based on big data capabilities was described. To achieve the accuracy and efficiency of customer service services from the perspective of technological innovation, how to use data analysis capabilities to build accurate customer support systems was explored. Finally, the important value of big data applied in customer operation scenarios and the prospect of customer operation were described.

Key words: internet of things, customer operation, big data



1 引言

随着物联网产业技术的日趋成熟,商业模式逐步形成,市场需求快速增长,物联网连接管理平台承载4万多个用户,数据量持续增加。每日新增60~70 GB接口数据和数以亿计的流量信息。

同时随着 Docker 容器技术、Kubernetes/Marathon 微服务框架等开源技术的迅速发展、逐渐成熟,传统的基于虚拟化的云架构受到了挑战。这些新技术代表了基于物理机的、应用于容器和大数据系统的下一代物联网大数据架构。

2 物联网客户运营的现状与挑战

物联网业务立足于做大连接服务规模,推动应用和部件服务突破,实现应用等非连接服务收入占比的提升,以实现快速提升物联网整体业务规模,是中国联通物联网面临的巨大技术挑战,归纳起来有以下几点。

- 现有技术掌握在部分第三方厂商手中,难以针对现有物联网数据进行更为灵活的大数据能力开放。
- 现有指标加工后无法直观地产生市场价值;系统对指标进行二次个性化加工或缺少二次加工;缺少可视化数据展示、用户画像及数据建模功能。
- 现有物联网大数据分析能力与物联网业务的快速发展之间不充分、不均衡。

另一方面,中国联通物联网的市场业务布局也面临着以下挑战。

- 在产业链方面的参与程度以及资源整合程度落后于友商。
- 在用户规模上还有很大的上升空间。

3 国内物联网行业现状

国内物联网产业生态构建进入了关键期,产

业各方围绕物联网平台加速布局,产业生态构建成为物联网产业竞争的焦点。当前,物联网产业已度过以企业自由生长为主的早期阶段,产业兼并整合加剧,物联网平台成为兼并抢夺的热点,巨头企业围绕产业生态主导权展开竞争,加速推动产业整合。产业价值和增长空间向软件和基于数据的服务转移。行业的产品、企业流程、生产方式、创新能力、商业模式和产业生态正在被软件和基于数据的服务定义。据麦肯锡预测,至2025年,软件商和集成服务商在物联网价值链中的份额将会进一步增加,达到60%~85%。物联网平台成为拉动终端与应用的控制点,产业纷纷率先布局平台。物联网平台凭借其对产业链上下游企业整合、促进应用发展的关键作用,成为产业生态构建的核心要素。

4 物联网客户运营平台的构建

中国联通研究院通过大数据技术及人工智能替代原有的线下物联网数据分析手段,提升物联网客户运营能力,实现客户运营服务的精准化、高效化,拟建设物联网客户运营平台,通过打造数据分析能力,构建精准的客户运营支撑体系。

物联网客户运营平台采用五层架构,从上到下分为展现层、应用层、服务能力层、数据层和系统层,如图1所示。其中,展现层为面向物联网运营者及管理者的Web页面,通过丰富的可视化指标展示,指导业务线开展运营分析工作,实时掌握物联网业务发展的脉络;应用层包含众多物联网大数据分析能力;服务能力层则包含大数据能力、负载均衡能力、镜像仓库能力、消息队列能力、日志分析能力和人工智能分析能力;第四层是数据层,包括关系型的数据库及分布式的文件系统;第五层是底层系统层,包括硬件系统和软件系统。

4.1 通过多维数据分析引擎实现模型加速

多维数据分析引擎(如图2所示)主要用于

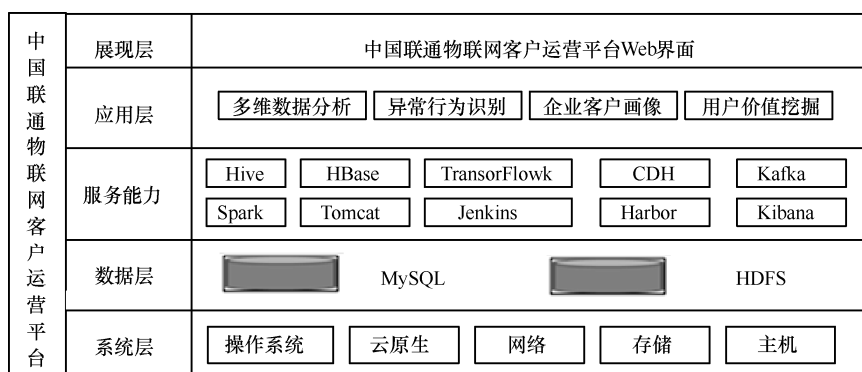


图1 中国联通客户运营平台技术架构

支持大数据生态圈的数据分析业务,解决大数据系统中TB级别数据的数据分析需求。多维数据分析引擎利用Hadoop的MapReduce框架实现,在数据ETL导入OLAP引擎时提前计算各维度的聚合结果,将用户设定的多维立方体缓存到HBase并持久化保存,通过预计算的方式缓存所有需要查询的数据结果。可根据数据仓库的实际配置实现跨集群的数据源抽取计算。具备交互式查询能力,与原有BI工具无缝整合。应用场景覆盖了销售、市场、管理等商务报表、预算决算、经济报表等,目前支持物联网业务报表160项指标的预设定模板查询、用户自定义指标查询以及秒级的快速查询服务。

提供REST数据接口服务等一系列工作,其架构如图3所示。

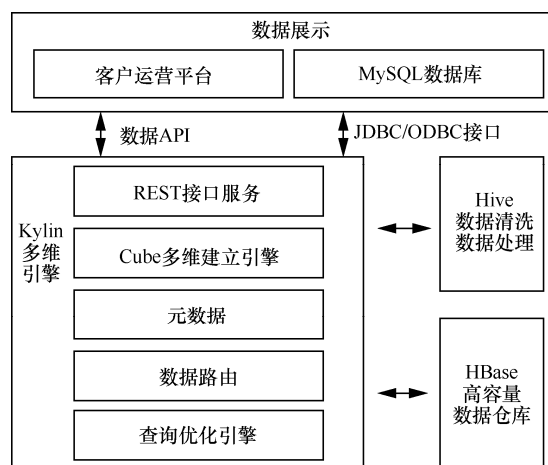


图3 OLAP技术架构

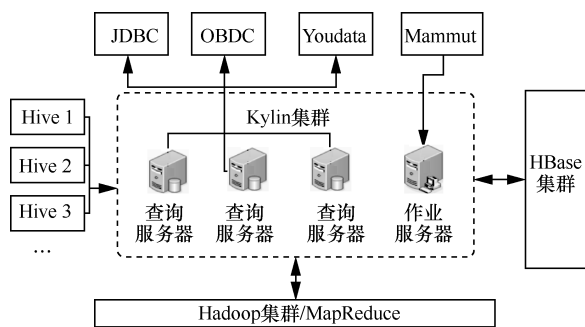


图2 多维数据分析引擎

中国联通基于技术先进性并结合现有Hadoop集群较为成熟的技术架构,采用麒麟技术开源技术作为多维数据分析引擎。在系统的实践中使用麒麟完成了从数据源抓取数据、使用ETL工具清洗数据、将数据转存到自己的存储引擎、

多维计算能力主要体现在OLAP Cube的计算上。一个Cube包含多个Cuboid,而Cuboid又是由原始数据聚合的数据。所以创建Cube可以看作在原始数据导入时做的一个预计算、预处理的过程。Kylin的强大之处在于充分利用了Hadoop的MapReduce并行处理的能力,高效处理导入的数据。

Kylin的数据来自于Hive,并作为一个Hive的加速器,希望最终的查询SQL类似于直接在Hive上查询。因此Kylin在建立Cube时需要从Hive获取Hive表的元数据。虽然有建立Cube的过程,但是并不想对普通的查询用户暴露Cube的存在。



首先根据 Cube 定义的事实表以及维度表,利用 Hive 创建一张宽表,抽取事实表上的维度的 distinct 值,将事实表上的维度以字典树方式压缩编码成目录,将维度表以字典树的方式编码,然后通过 Hadoop 的 MapReduce 把上一步中的宽表文件作为输入,创建 N -Dimension cuboid,最后重复以上步骤,根据前一次的结果串行生成下一次的输入文件。形成 n 维结果;另一方面,根据生成的 Cuboid 数据量计算 HTable 的 Region 分割策略,创建 HTable,将 HFile 导入多维引擎。

4.2 基于人工智能的物联网异常行为识别

基于人工智能的异常行为监控模型如图 4 所示,对实时采集的物联网海量终端数据的异常行为进行监控与检测。基于现有 Jasper 平台 Hadoop 集群上存储的物联网数据,例如少部分被标记为行为异常的数据和海量未标记的数据,使用 Spark 数据加工流程提取出重要字段对应的关键数据,采用基于正样本和未标记样本的半监督分类模型,省去大量标记数据的人力与物力。利用已标记的少量异常行为数据和大量未标记的数据,通过二阶段算法模型,第一阶段使用朴素贝叶斯算法获取可靠的负样本数据,第二阶段利用随机生成树、支持向量积、逻辑回归、 K 近邻分类等分类模型进行数据分类的训练,可获得甄别异常行为的模型参数,从而对实时采集的物联网终端数据做出及时判断,并通过数据库中的已有数据信

息按年月日获取物联网用户异常行为总结。通过人工智能的异常行为监控模型,可以帮助业务人员及时发现用户的异常行为。

机器学习可以分为监督式学习、半监督式学习、无监督式学习以及强化学习。根据物联网部门提供的数据特征,即已知少部分用户异常行为数据,未标注的数据中既有用户异常行为数据又有用户行为数据,这使得监督学习不可行,所以需要通过建立基于正样本和未标记负样本的半监督学习模型来预测物联网用户异常行为群。

半监督学习使用大量的未标记数据以及少量标记数据来进行建模。在现实任务中,未标记样本多、有标记样本少是一个普遍现象,利用好未标记样本来提升模型泛化能力,就是半监督学习的优势所在。半监督学习不仅可以极大地减少人为的工作量,同时又可以获得准确性较高的结果。在物联网用户异常行为识别的案例中,有大量的未标记数据以及少部分确定的用户异常行为数据,这种数据特征非常适合使用正样本和未标记负样本的半监督学习方法进行建模。在物联网用户异常行为识别的案例中,有大量的未标记数据以及少部分确定的用户异常行为数据,这种数据特征非常适合使用随机森林及 SVM 的算法进行结果预测,下面通过这两种方法来进行用户异常行为的识别。

4.2.1 随机森林

集成学习 (ensemble learning) 是通过构建并

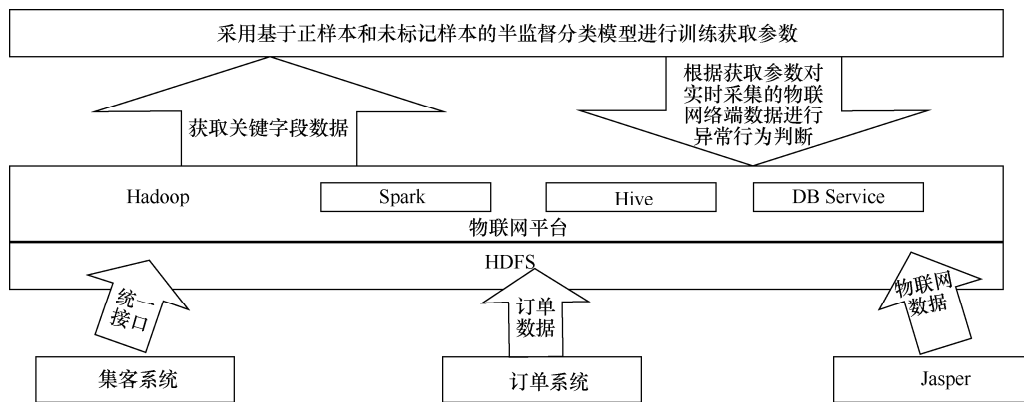


图 4 人工智能异常行为监控模型

结合多个学习器来完成学习任务, 主要包含两类, 一类是个体学习器间存在强依赖关系、必须串行生成的序列化方法, 比如提升树 GBDT、XGBoost 等, 主要思想是 Boosting 迭代将弱学习器提升为强学习器; 第二类是个体学习器之间不存在强依赖关系、可同时生成的并行化方法, 该类的代表是 Bagging。随机森林是 Bagging 的一个拓展变体。

随机森林在以决策树为基础学习器构建 Bagging 集成的基础上, 进一步在决策树的训练过程中引入了随机属性选择。传统决策树在选择划分属性时是在当前节点的属性集合(假定 m 个属性)中选择一个最优属性; 而在随机森林中, 对决策树的每个节点, 先从该节点的属性集合中随机选择一个包含 k 个属性的子集, 然后再从这个子集中选择一个最优属性用于划分。这里的参数 k 控制了随机性的引入程度: 若令 $k=m$, 则决策树的构建与传统决策树相同; 若令 $k=1$, 则随机选择一个属性用于划分, 一般情况下 $k = \text{lb}m$ 。

随机森林还加入了随机属性扰动, 这使得最终集成的泛化性能可通过个体学习器之间差异度的增加而进一步提升。随机森林的收敛性与 Bagging 相似, 随机森林的起始性能相对较差, 特别是只包含一个基学习器时(因为通过引入属性扰动, 随机森林中个体学习器的性能往往有所降低), 然而随着个体学习器数目的增加, 随机森林通常会收敛到更低的泛化误差。

在随机森林中将生成很多的决策树, 当基于某些属性对一个新的对象进行分类判别时, 随机森林中的每一棵树都会给出自己的分类选择, 并由此进行“投票”, 森林整体的输出结果将会是票数最多的分类选项, 计算式如下:

$$H(x) = \max_Y \sum_{i=1}^k I(h_i(x) = Y) \quad (1)$$

其中, $H(x)$ 表示随机森林分类结果, $h_i(x)$ 是单个决策树分类结果, Y 表示分类目标, $I(\cdot)$ 为示性函

数, 使用分类投票策略, 根据投票结果确定分类, 相关模型的详细推导见参考文献[13]。

支持向量机(support vector machine)是形式为 $f(x) = w^T x + b$ 的线性方程组, $w^T x$ 是权向量和输入向量 x 的内积。SVM 用作分类器, 如果 $f(x) > 0$, 则分为 1 类, 否则分为 -1 类。SVM 的核心思想是选出一个超平面将正负样本分开, 同时最大化最小边界。 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ 是一组训练数据, x_i 代表输入向量, y_i 代表分类标签, $y_i \in \{1, -1\}$ 。所以寻找超平面的问题变成了以下问题:

$$\min \frac{1}{2} w^T w, \quad y_i(w^T x_i + b) \geq 1, i=1, 2, \dots, n \quad (2)$$

为了处理一些因为噪音数据而造成的没有超平面的情况, 提出了 Soft-margin SVM 来寻找超平面的方法, 计算式如下:

$$\min \left(\frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \right), \\ y_i(w^T x_i + b) \geq 1 - \xi_i, i=1, 2, \dots, n, \quad C \geq 0 \quad (3)$$

其中, C 可以控制训练错误的大小。

4.2.2 Biased SVM

在处理不平衡数据分类时, 支持向量机遇到的主要问题是分类面会靠近少数类, 从而会分错少数类, 所以提出了一种新的策略, 将正负样本的权重分开对待, 对于 PU 学习问题, 要尽量将正样本分开, 因此正样本的权重系数 C 会比较大。以下是一个效果比较好的用于第二阶段的算法, 由 SVM 模型演化而来, 求解目标函数为:

$$\min \frac{1}{2} w^T w + C_+ \sum_{i=1 \dots k} \varepsilon_i + C_- \sum_{j=k+1 \dots n} \varepsilon_j, \\ y_i(w^T x + b) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0 \quad (4)$$

这种将正负样本分开对待的方式使得正确率提升很多。

4.3 运用 ETL 进行大数据企业用户画像分析

数据源的数据通过数据建模的定向处理, 按照特定的规则通过 ETL 的处理后, 在数据仓库中



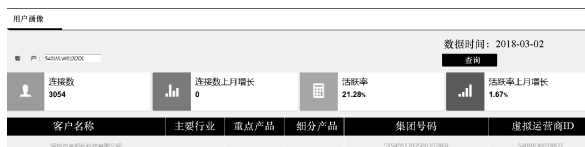
进行多维度分析。将本身拥有大量用户数据的数据平台和可视化数据工具连接起来,根据不同的用户交互场景,挖掘数据平台的价值,系统用户能够根据需要,随时自主地分析不同用户特征,快速洞察用户需求。

从数据源到最终企业用户画像展现分成如下几层。

- 数据源: 包括来自各个业务系统的数据源,其载体包括数据库、文件、大数据平台等。
- 数据建模: 根据用户画像建模体系配置数据模型。
- ETL: 将数据从来源端经过抽取(extract)、交互转换(transform)、加载(load)至数据集市。每个数据集市是基于一个主题做好轻量建模的细节数据,数据按照各自的存储方式,被高效压缩、打好标签。当需要计算时,采用分布式计算来进行数据计算,最终将结果送往用户画像可视化分析层做展现。
- 用户画像可视化分析: 采用开源 echarts 作为前端交互组件。用户可以通过主流浏览器来访问用户画像系统,用户还可通过移动终端来访问系统。
- 雷达图、用户标签体系及建议系统: 可以根据特定用户的不同纬度数据指标,根据数据指标具体数值,自动对应用户标签及给定特定的维系建议。

用户画像功能共有 3 个部分: 用户画像 4 个参数与用户信息显示部分、用户画像雷达图、用

户画像 3 个柱图部分。通过用户画像页面的客户 ID 查询,可以动态显示 4 个参数与用户信息部分以及客户具体的激活率趋势、出入账情况趋势及订购情况趋势。界面展示如图 5 所示。



用户画像

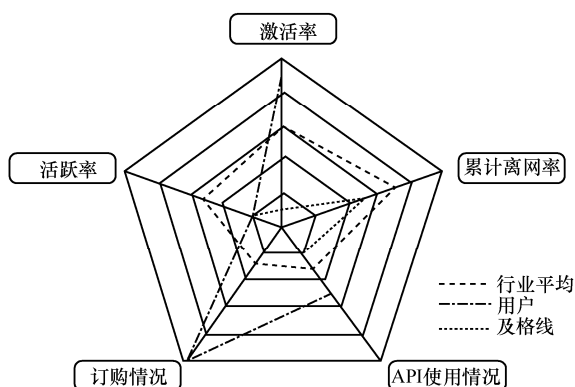


图 5 界面展示

用户画像默认显示: 用户在载入用户画像功能后,会对登录的用户身份进行验证,根据不同登录用户的身份,筛选对应的可视化用户画像的数据,将数据封装为特定的用户画像可视化的格式并返回前端进行显示,显示的内容包括上述 3 部分内容。用户画像默认显示流程如图 6 所示。

用户画像查询功能: 当需要对特定用户进行查询时,可以通过用户画像页面的用户 ID 查询。后台通过筛选后,将数据封装为特定的用户画像可视化的格式,返回前端进行显示,显示的内容包括上述 3 部分内容。用户画像查询流程如图 7 所示。

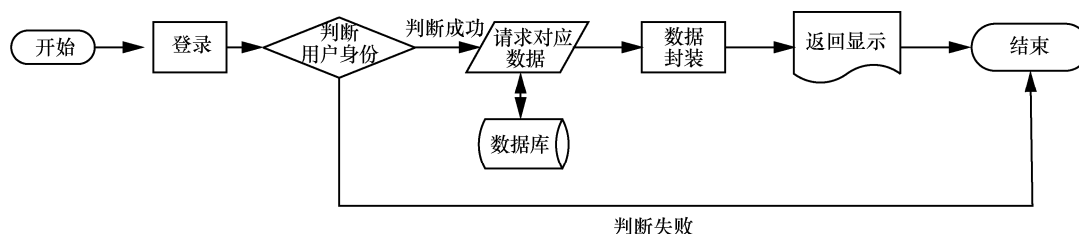


图 6 用户画像默认显示流程

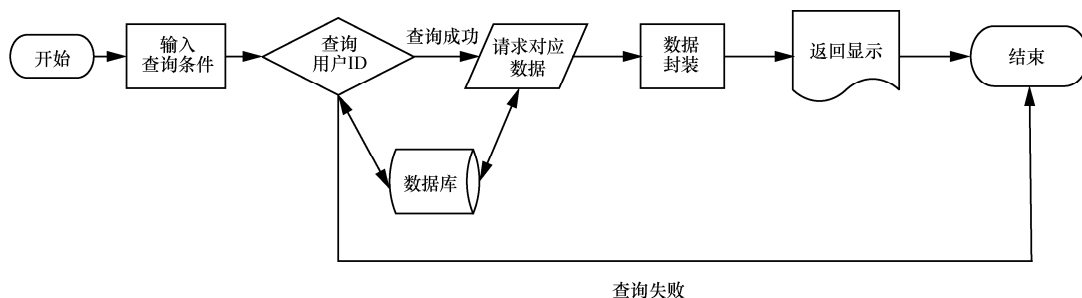


图7 用户画像查询流程

4.4 虚拟化容器云技术整合硬件资源

虚拟化容器云技术串联了物联网应用和底层的硬件设备，向下对接 IaaS 平台，可以完成物理资源的粗颗粒度调度、大数据资源的一体化调度，向上提供大数据微服务化能力，容器化承载各 IT 域无状态应用的微服务能力，完成 IT 资源的集中管理。在同一个容器云集群上部署离线计算与分析、定时批量处理和实时流式计算等不同类型的大数据应用，通过在资源池化的集群中混合部署 Hadoop、Spark、Kafka、HBase、Storm、Flink 等大数据应用，将不同类型的应用自动混合部署到整个集群，不受单个服务器边界限制，这样提升了资源利用率。业务运营和后台管理全部采用可

视化操作，纵向提供能力、资源、监控、集群和租户管理，横向提供节点、容器以及框架的全生命周期管理，整体上完成从资源、应用到服务的一体化运营。

微服务能力管理如图 8 所示，客户运营分析平台及相关组件开发过程中借助容器技术、微服务架构、持续集成工具，让物联网项目可以敏捷开发、一键部署。通过不断地迭代来满足物联网用户快速更新的需求，自动化测试保障了代码质量，高效的代码和流程管理工具保障了代码管理、打包、发布流程可视化，从而使物联网项目开发团队高效的自我管理成为可能。

功能架构如图 9 所示，系统功能基于 Jenkins

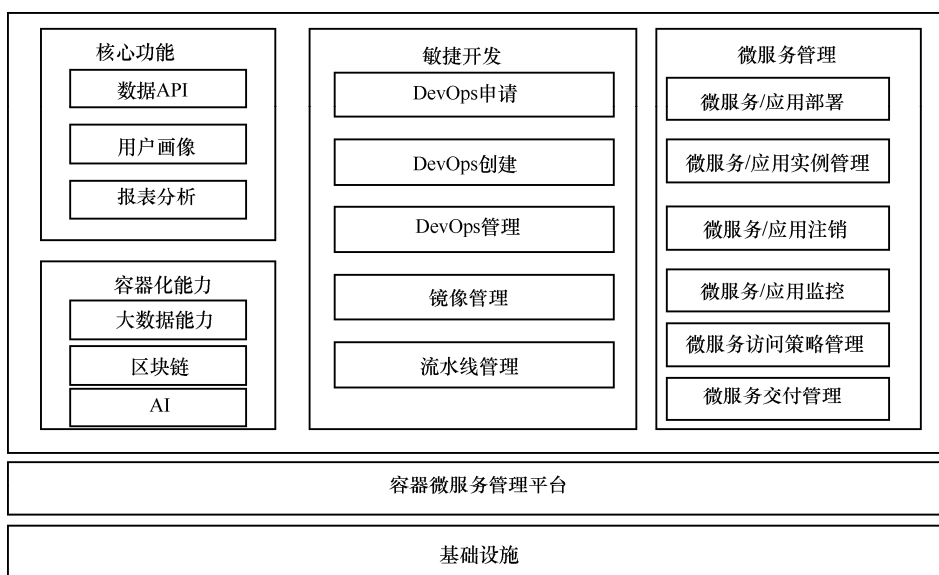


图8 微服务能力管理



执行构建，Jenkins 项目触发方式有 4 种：手动触发、代码库更新触发、定时触发和上游触发。当 Jenkins 执行任务时，首先调用接口，通知系统执行构建任务。Jenkins 通过在系统中构建配置信息，查询需要执行的测试任务，确定调用方式后，再调用具体的测试服务。测试服务执行完成后，将结果返回给 Jenkins，Jenkins 存储结果后，结束执行。

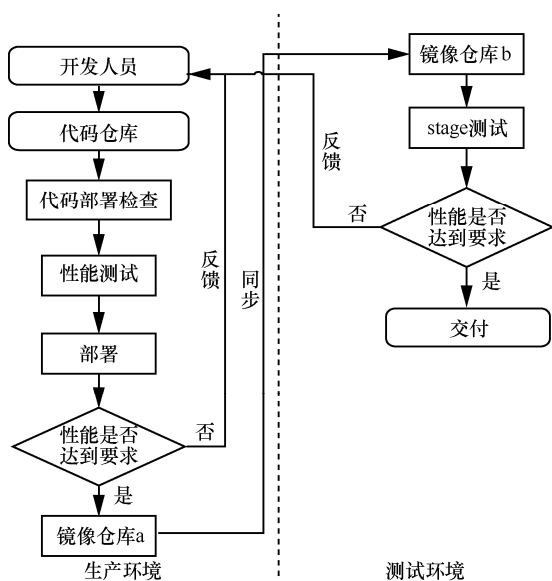


图 9 功能架构

在技术方面，系统使用 HTML 实现，采用 Bootstrap+SpringBoot 框架，以关系型数据库 MySQL 作为持久化存储方案。

集成是软件开发过程中的一个重要环节，如果不能早集成、常集成，那么等到最后再集成会出现很多问题，给项目带来不可预期的麻烦，为了提升开发的效率、降低项目的不可预见性，实践中频繁地进行集成，每一次集成都通过自动化构建来验证，从而尽快地发现集成错误。持续集

成里的构建不仅仅包含编译、链接、组成可执行程序，还应包括对这个可执行程序进行的自动化测试。开发人员被要求在这个自动化过程中的每一步都不能出错，而最重要的一步就是自动化测试，可以这么认为：自动化测试是持续集成的核心。

作业提交流程如图 10 所示，用户可以上传 war 包类程序包，通过 war 的编译、压缩上传，打包形成 img 类文件，或者直接上传 img 类文件。系统将上传的文件归档形成，可以启动的容器云资源能力，以使用户在需要的场景下启动能力。

5 物联网客户运营业务上的技术创新实践

（1）高速数据处理能力

针对物联网客户运营可利用目前业界最流行的开源大数据和人工智能技术进行研发，如 Spark、TensorFlow 等，在工具上使用了 ETL 引擎、多维数据分析引擎以及数据网关。在数据的收集、存储、加工、融合的全流程管理中，利用开源技术+工具+平台的模式形成自动化数据生产线，大幅提高数据开发效率，降低研发人工成本。

（2）多元多业务线数据融合

物联网数据与以往电信业务相比，数据量更为庞大，这就要求业务系统必须融合物联网连接平台的详单数据、中间数据和结果数据等包含物联网连接业务的全生命周期数据。但就数据价值而言，相较于同行业其他平台，运营商具备自身优势，可将数据融合度最大化。可从位置、连接、运营、网元等多维度共同优化运营场景，丰富数据开放维度，提升物联网服务

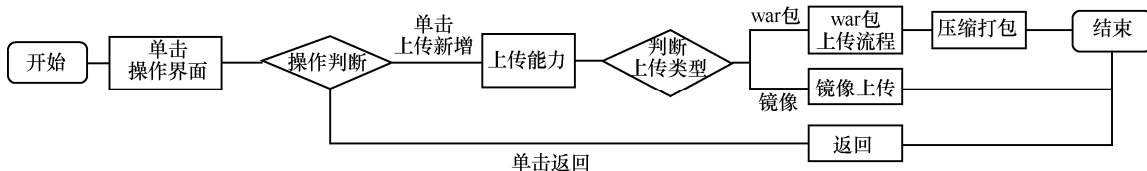


图 10 作业提交流程

附加价值。

(3) 虚拟化容器技术

要想承载物联网庞大的客户运营数据, 必须要求系统具备大数据开放能力和应用云化的能力, 形成了从主机、云、大数据、人工智能到应用的一体化解决方案。应用与平台解耦合, 这样才能发挥平台化最大效能和优势, 在实现资源集约的同时保证效率最大化。

6 结果验证

物联网客户运营分析需通过测试与验证, 通过发现问题来改进系统中可能存在的问题, 最终实现项目的完善。本章将对系统的性能和功能两大方面进行系统性的测试与验证。

本文以 CentOS 7 为验证环境, 验证了主机

172.18.50.62-82 和 20 台 x86 机架服务器。

主机部署性能要求见表 1。

软件环境包含服务器软件环境和测试工具。

具体的配置见表 2。

通过系统白盒测试进行功能性检验, 测试人员利用程序内部的逻辑结构及有关信息, 设计或选择测试用例, 对程序所有逻辑路径进行测试。通过在不同点检查程序状态, 确定实际状态是否与预期的状态一致。因此, 白盒测试又称为结构测试或逻辑驱动测试。

白盒测试是穷举路径测试。在使用这一方案时, 必须检查程序的内部结构, 从检查程序的逻辑着手, 得出测试数据。

白盒测试主要是对程序模块进行如下检查。

- 对程序模块的所有独立的执行路径至少测

表 1 服务器硬件环境

测试环境			
名称	配置	与实际差异	备注
机架式服务器	E5-2620v4 8core×2 128 GB (8×16 GB、4×32 GB) 系统盘: 240 GB s3 510 以上 SSD 硬盘: 4 TB (6×12) 网卡: 10GE×4GE×2GE×1 (IPMI) RAID 卡	模拟	实验室环境性能弱于实际环境
数据库系统	MySQL 5.7 版本, 容器化应用, 单例 16 核 CPU, 32 GB 内存, 持久化存储可伸缩	模拟	实验室环境性能弱于实际环境
负载均衡器	Nginx	模拟	
光纤交换机	-	无法模拟	

表 2 服务器软件环境

测试环境			
软件名称	版本	与实际差异	备注
操作系统	CentOS 7	一致	
容器运行系统	Docker 1.11	一致	
容器编排系统	Kubernetes 1.8	一致	
框架编排系统	Open DC/OS 1.05, 内置 Marathon	一致	
集群编排系统	Open DC/OS 1.05	一致	
资源编排系统	Mesos 1.0.0	一致	
数据库系统	MySQL 5.7	一致	



试一遍。

- 对所有的逻辑判定，取“真”与取“假”的两种情况都至少测一遍。
- 在循环的边界和运行的界限内执行循环体。
- 测试内部数据结构的有效性。

本节通过 4 个测试用例，阐述电信运营商物联网客户运营分析系统的测试过程。测试用例覆盖多维搜索引擎性能测试、用户画像流程测试、基本功能测试、虚拟化容器能力测试四大方面。

6.1 测试 1: 经营多维分析与 Oracle 性能对比测试

经营多维分析与 Oracle 性能对比测试，对比经营多维分析所使用的多维数据分析引擎和传统 Oracle 的性能。

测试业务场景：用户可通过经营多维分析页面得知具体某一天的详细指标及趋势分析图。

测试预设条件：登录用户需具有物联网运营平台查看权限。

测试步骤如下。

- 单击<经营多维分析>，进入经营多维分析页面，在上方日期中选择具体日期。Oracle 和多维引擎各执行一次。测试检查点为当前页面中展示当天数据的分析数据及趋势图。如果正常显示，且数据与数据库保持一致证明功能正常。
- 单击<应收数据><连接数><客户统计>按钮，进入数据分析页面。Oracle 和多维引

擎各执行一次。测试检查点为记录各功能前台展示的响应时间。

预期结果：经营多维分析页面展示流程执行正常，显示进度节点正常。

测试结果：测试结果通过，测试截图如图 11 所示。

- 通过使用麒麟引擎与 Oracle 进行比较，当查询数据量较小时，Oracle 和麒麟引擎的差距不大。麒麟引擎并未体现出预先计算加载的优势。
- 当执行数据量不大，但表聚合、关联查询较为复杂的连接数聚合查询时，Oracle 相比麒麟引擎的预先计算功能体现出较大的响应时间差距。
- 当执行数据量大、表聚合不多的客户统计查询时，麒麟引擎依靠 HBase 组件优异的查询性能，大大减少了查询的响应时间，测试结果如图 12 所示。

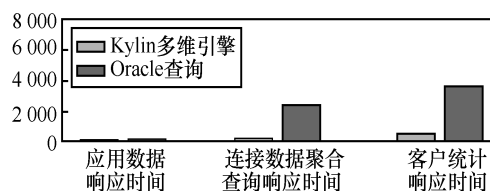


图 12 查询响应测试结果

6.2 测试 2: 运营多维分析功能测试

运营多维分析是面向业务线运维人员的功能



图 11 经营多维分析界面功能演示

页面,主要测试运营多维分析页面展示是否正常,功能使用是否正常。

测试业务场景:用户可通过运营多维分析功能页面查询不同省份在不同年份、月份的连接数及出账情况。

测试预设条件:登录用户需具有物联网运营平台查看权限。

测试步骤:使用管理员账号登录物联网平台客户运营系统首页,选择要查找的省份、年份、月份,单击运营多维分析操作中<查询>按钮,查看省份在不同年份、月份的连接数及出账情况。测试检查点为弹出分析图页面,检查流程节点显示正常。

预期结果:运营多维分析页面展示流程执行正常,流程顺序正确,显示进度节点正常。

测试结果:测试结果为通过,展示功能全部正常,图标数据响应时间为 158 ms。

运营多维分析功能页面如图 13 所示。

6.3 测试 3: 用户画像功能测试

用户画像功能面向业务线用户,帮助用户快速描述客户特征,并给出智能运营建议。本项测试主要测试用户画像查看功能的可用性。

测试业务场景:通过物联网平台客户运营系

统,查看所有已有客户的客户特征,并通过界面展示。

测试预设条件:登录用户需具有物联网运营平台查看权限。

测试步骤:使用管理员账号登录物联网平台客户运营系统首页,在页面上方<客户>中输入客户的虚拟 ID—<查询>。检查用户画像查看页面是否正常显示、对应的客户五大特征是否正常显示以及用户基本资料是否与数据库一致。

预期结果:查看各类型客户特征数据正常显示,统计信息无误,导出数据正常。

测试结果:用户可以通过主流浏览器来访问用户画像系统,用户通过移动终端来访问系统。

雷达图、用户标签体系及建议系统,可以根据特定用户的不同纬度数据指标,根据数据指标具体数值,自动对应用户标签及给定特定的维系建议。响应时间为 87 ms。

用户画像特征显示及运营建议页面如图 14 所示。

6.4 测试 4: 虚拟化容器云应用运行性能测试

本项测试的用意为虚拟化容器云的应用承载性能。电信运营商物联网客户运营分析系统数据中包含海量的物联网数据,所接入的大数据能力



图 13 运营多维分析功能页面

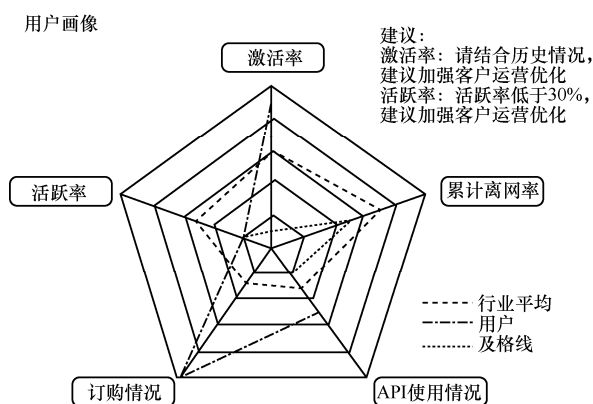


图 14 用户画像特征显示及运营建议页面

通过虚拟化容器的方式进行承载, 承载能力可随数据量线性增长。

测试业务场景如下。

- 启动 200 个数据处理容器同时并发, 处理流程结果数据。
- 查看系统相关的性能参数情况: CPU 使用率、磁盘空间的使用情况、内存的使用率、剩余内存百分比。

测试预设条件: 通过容器云平台配置 200 个数据处理容器的同时运行脚本。

测试步骤如下。

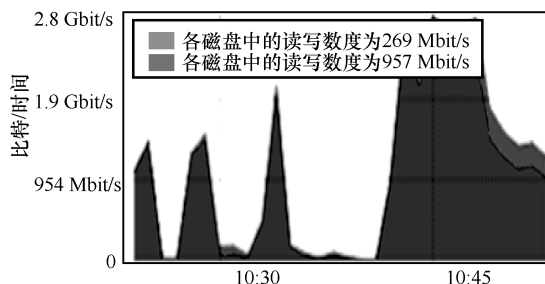
- 使用管理员账号登录物联网平台客户运营系统。
- 通过容器云技术同时启动 200 个数据处理容器, 并开始处理测试数据。同时检查系统的 CPU 使用情况、磁盘读写的情况。

预期结果: 查看各类型资源数据正常显示, 统计信息无误, 导出数据正常。

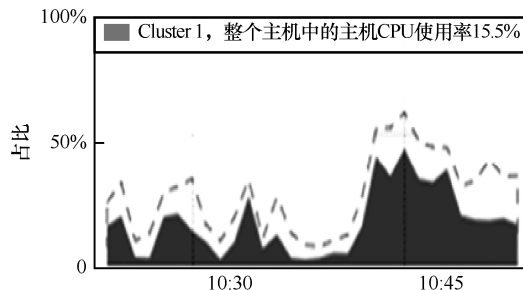
测试结果: 测试通过, 容器的资源使用及承载速度随数据量增加而线性增加。

测试结果分析如下。

- 测试开始时 200 个容器系统响应时间在 8s 以内。容器的资源占用情况随着数据量的起伏而起伏。
- CPU 使用、内存、读写均在可承载范围之内, 显示参数合理, 测试截图如图 15 所示。



(a) CPU使用情况



(b) 数据读写情况

图 15 容器云集群读写及 CPU 使用情况

7 结束语

物联网客户运营分析系统旨在通过容器技术、人工智能、多维数据引擎、大数据挖掘等创新技术的应用, 解决中国联通面临的实际物联网企业客户的运营难题。通过虚拟化容器技术实现应用的弹性伸缩, 形成在物联网业务下高负载的大数据服务能力。融合机器学习模型形成企业客户行为的指标监控。充分发挥物联网数据的高可见度、高真实性的数据特点, 深度挖掘物联网企业客户价值, 为物联网客户运营业务的开展构建坚实的技术基础。

参考文献:

- [1] BURNS B, GRANT B, OPPENHEIMER D, et al. Borg, Omega, and Kubernetes[J]. Communications of the ACM, 2016, 59(5): 50-57.
- [2] BAIER J. Getting started with Kubernetes[M]. New York: ACM Press, 2015.
- [3] BERNSTEIN D. Containers and cloud: from LXC to Docker to Kubernetes[J]. IEEE Cloud Computing, 2016, 1(3): 81-84.
- [4] 吴龙辉. Kubernetes 实战[M]. 北京: 电子工业出版社, 2016: 13-211.
- WU L H. Kubernetes combat[M]. Beijing: Electronic Industry

- Press, 2016: 13-211.
- [5] 丁飞. 物联网开放平台[M]. 北京: 电子工业出版社, 2018.
DING F. Internet of things open platform[M]. Beijing: Electronic Industry Press, 2018.
- [6] SMART J F. Jenkins 权威指南[M]. 郝树伟, 于振苓, 熊熠, 译. 北京: 电子工业出版社, 2016.
SMART J F. Jenkins: the definitive guide[M]. Translated by HAO S W, YU Z C, XIONG Y. Beijing: Electronic Industry Press, 2016.
- [7] KAKADIA D. Mesos 大数据资源调度与大规模容器运行最佳实践[M]. 崔婧雯, 刘梦馨, 译. 北京: 电子工业出版社, 2016.
KAKADIA D. Mesos big data resource scheduling and large-scale container operation best practices[M]. Translated by CUI J W, LIU M X. Beijing: Electronic Industry Press, 2016.
- [8] RYZA S, LASERSON U, OWEN S, et al. Spark 高级数据分析[M]. 龚少成, 译. 北京: 人民邮电出版社, 2017.
RYZA S, LASERSON U, OWEN S, et al. Advanced analytics with Spark[M]. Translated by GONG S C. Beijing: Posts&Telecom Press, 2017.
- [9] 余何. PaaS 实现与运维管理基于 Mesos+Docker+ELK 的实战指南[M]. 北京: 电子工业出版社, 2016.
YU H. PaaS implementation and operation and maintenance management based on Mesos+Docker+ELK practical guide[M]. Beijing: Electronic Industry Press, 2016.
- [10] ISMAIL B I, GOORTANI E M, KARIM M B A. Evaluation of Docker as edge computing platform[C]//2015 IEEE Conference on Open Systems (ICOS), Aug 24-26, 2015, Bandar Melaka, Malaysia. Piscataway: IEEE Press, 2015: 130-135.
- [11] TAKASE W, NAKAMUR T, WATASE Y, et al. A solution for secure use of Kibana and Elasticsearch in multi-user environment[Z]. 2017.
- [12] 苏敏坚. 大数据是运营商复兴的“起搏器”[J]. 中国电信业, 2017(4): 40-41.
SU M J. Big data is the ‘pacemaker’ for operators to revitalize[J]. China Telecommunication Trade, 2017(4): 40-41.
- [13] 范铭. 运营商进入调速换挡关键时期大数据创造新机遇[J]. 通信世界, 2017(16): 54.
FAN M. Operators enter new era in the key period of speed control shifting, creating new opportunities[J]. Communications World, 2017(16): 54.
- [14] 云晴. 西班牙电信大数据实践启示“数据可视化”至关重要[J]. 通信世界, 2017(13).
YUN Q. Telefónica’s big data practice enlightens ‘data visualization’ is essential[J]. Communications World, 2017(13).
- [15] 李莺. 基于大数据分析的运营商主动服务研究[J]. 移动通信, 2016(24): 87-90.
LI Y. Research on active services for operators based on big data analysis[J]. Mobile Communications, 2016(24): 87-90.
- [16] 穆致君, 何阳, 杨丹. 运营商的大数据业务发展模式[J]. 现代电信科技, 2016, 46(1): 8-12.
MU Z J, HE Y, YANG D. The analysis of enterprise patters of big data of telecom operators[J]. Modern Science & Technology of Telecommunications, 2016, 46(1): 8-12.
- [17] 费嘉琪. 运营商发展大数据的机遇和挑战[J]. 数字通信世界, 2017(8): 115.
FEI J Q. Opportunities and challenges of big data development in operators[J]. Digital communication World, 2017(8): 115.
- [18] 王晓霞, 刘静沙, 许丹丹. 运营商大数据用户画像实践[J]. 电信科学, 2018, 34(5): 127-133.
WANG X X, LIU J S, XU D D. A practice of user profile in big data[J]. Telecommunications Science, 2018, 34(5): 127-133.
- [19] 张基恒, 李大中, 张呈宇, 等. DC/OS 关键技术与应用场景[J]. 电信科学, 2016, 32(12): 13-19.
ZHANG J H, LI D Z, ZHANG C Y, et al. Key technologies and application scenario of DC/OS[J]. Telecommunications Science, 2016, 32(12): 13-19.
- [20] 李董, 魏进武. 区块链技术原理、应用领域与挑战[J]. 电信科学, 2016, 32(12): 20-25.
LI D, WEI J W. Theory, application fields and challenge of the blockchain technology[J]. Telecommunications Science, 2016, 32(12): 20-25.

[作者简介]



张勋 (1990-), 男, 中国联合网络通信有限公司研究院大数据研究中心软件开发工程师, 主要研究方向为开源技术、容器技术。



翟春辉 (1987-), 男, 中国联合网络通信有限公司研究院大数据研究中心软件开发工程师, 主要研究方向为开源技术、应用技术和区块链技术。

张呈宇 (1988-), 男, 中国联合网络通信有限公司研究院大数据研究中心平台技术研究开发室主任, 主要研究方向为大数据、云计算、容器技术等。

贾捷 (1992-), 男, 中国联合网络通信有限公司研究院大数据研究中心, 主要研究方向为大数据、云计算、容器技术等。

龙岳 (1978-), 女, 中国联合网络通信有限公司研究院大数据研究中心, 主要研究方向为大数据框架新技术研究和开发。