

【本文献信息】王东明.基于Hadoop技术的广电增值业务生命周期评价模型研究与应用[J].广播与电视技术,2023,Vol.50(7).

基于Hadoop技术的广电增值业务 生命周期评价模型研究与应用

王东明

(江苏省广电有线信息网络股份有限公司, 江苏 210000)

【摘要】本文阐述了广电增值业务的生命周期评价模型以及应用,介绍了业务探针数据的采集、入库、处理等相关概念。同时针对各个业务的用户覆盖数、点播量、点播用户、销账额,建立了业务生命周期评价模型、业务贡献度评价模型,用于对这些增值业务的评价提供数据参考。

【关键词】Hadoop, 探针, 业务生命周期评价, 业务贡献度评价

【中图分类号】TP311.1 【文献标识码】B 【DOI编码】10.16171/j.cnki.rtbe.20230007013

Research and Application of Life-cycle Evaluation Model of Broadcasting Value-added Services Based on Hadoop Technology

Wang Dongming

(Jiangsu Broadcasting Cable Information Network Co., Ltd., Jiangsu 210000, China)

Abstract This paper expounds the life-cycle evaluation model and application of broadcasting value-added services, and introduces related concepts such as collection, storage, and processing of service probe data. At the same time, according to the user coverage, viewing times, viewing users, and write-off amount of each business, a business life-cycle evaluation model and a business contribution evaluation model are established to provide data reference for the evaluation of these value-added services.

Keywords Hadoop, Probe, Business life-cycle evaluation, Business contribution evaluation

0 引言

广电增值业务作为一种产品,其用户的订购数以及盈利能力难以一成不变,会随着时间的推移而改变。这其中伴随着产品的新生、生长、成熟和衰落的过程,类似生物的生命历程,所以被称之为业务生命周期。广电业务的生命周期是指产品从进入市场到下架的全过程,开始和结束分别以进入和退出市场为标志。

面对日趋激烈的市场竞争,江苏有线引入了众多互联网形态的视频点播专区作为自身业务的补充,本文对如何考察与衡量这些增值业务的落地效果,以及它们的生命周期进行了探索与研究。通过业务的覆盖用户数、业务访问量、点播量、销账额等指标形成一套评价体系,实现业务价值的精准评价,

从而找出自身业务与合作伙伴提供的增值业务之间的资源差异,为公司购买影视剧版权提供决策参考,客观评价这些增值业务的落地效果。

本文拟挑选江苏有线某地级市的电影院线等15个增值业务,从业务周期评价、业务贡献度评价两个角度对其进行评估。

1 主要相关技术

1.1 Hadoop相关技术

Hadoop是一个分布式处理框架。主要核心项目由HDFS和MapReduce组成。HDFS(Hadoop Distributed File System)全称为分布式文件系统,HDFS由四部分组成,即HDFS Client、NameNode、DataNode和Secondary NameNode。它是基于流数据模式访问和处理超大文件的需求而开发的,是分

表1 探针数据采集标准

参数名称	字段	类型
终端设备编码	uservalue	String
终端区域标识	DeviceRegionId	Uint
操作时间	Time	String
业务类型	ServiceType	Uint
门户编码	Portalcode	string
节目编号	ProgramID	String
点播系统编号	SystemID	String
节目名称	ProgramName	String
栏目编号	CategoryId	String
栏目名称	CategoryName	String
播放类型	PlayType	Uint
按钮事件	btnType	String
探针记录	Result	String
操作类型	VODOperateType	Uint
播放速率	Pace	Uint
进入方式	EnterType	Uint

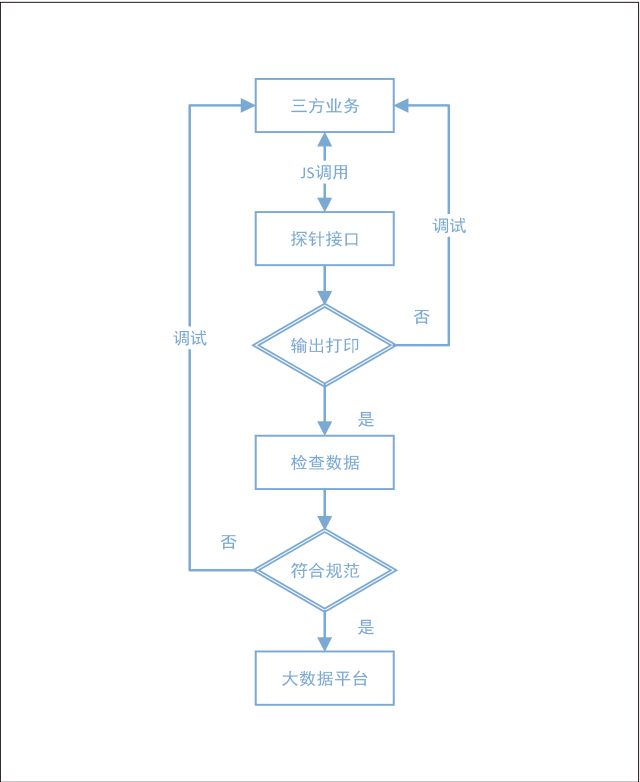


图1 探针采集工作流程

布式计算中数据存储管理的基础；它可以运行在比较廉价的服务器上，为大规模数据处理降低了成本；它具有的高容错、高可扩展性、高可靠性、高吞吐率等特征，可以为海量数据的存储提供安全保障。MapReduce 是面向大数据并行处理的计算模型、框架和平台，主要有以下优点：

1. MapReduce 作为一种分布式计算模型，它将大量的数据先拆分，然后分发到多台计算机中，这些数据可以在不同计算机同时进行计算。
2. 在处理数据的效率方面，相比于单节点的数据处理方式，MapReduce 是在各个计算机中同时进行的，因而节省了大量时间。
3. MapReduce 处理效率的优势也得益于自身的本地计算原理，且操作起来简单，只需要用户实现 map 和 reduce 函数，剩下的工作都可以由框架本身去处理完成，有效地减少了为计算做准备期间的数据传输。

1.2 Hive相关技术

Hive 是 Hadoop 的一个数据仓库工具，是将结构化的数据文件映射为一张数据库表，并提供简单的 SQL 查询功能，可以将 HiveQL 语句转换为 MapReduce 任务运行。它提供了一系列的工具，可以用来进行数据提取转化加载（ETL）。Hive 没有特定的数据格式。Hive 可以很好地工作在 Thrift 之上，控制分隔符，也允许用户指定数据格式。

1.3 探针数据采集相关技术

通过制定规范的探针采集规范，形成探针采集规范接口 API，然后在业务应用代码里以埋点的方式调用规范接口 API 来获取用户的行为数据，实现探针数据的收集以供大数据分析使用。

2 数据平台设计

数据平台主要包含以下流程：

探针数据标准制定 -> 探针数据布点采集 -> 探针数据入库 -> 探针数据清洗与统计。

2.1 探针采集标准制定

探针采集标准关系到采集数据的完整性、可分析性，基于有线的用户行为数据特征，我们制定了探针数据采集标准，如表 1 所示。

2.2 探针采集技术路线

前端采用 Ajax 异步请求后台探针接口，避免服务端出现问题导致页面加载缓慢。服务端采用 Redis 缓存存储临时数据，再通过每分钟一次的定时任务将探针数据写入硬盘，以应对高并发带来的 I/O 压力，工作流程如图 1 所示。

2.3 探针数据入库

基于 DataOS 平台搭建 ETL 流程, 对采集的数据入 HDFS 库。ETL 流程包含的节点如图 2 所示。

执行端命令节点是从探针服务器取数到本地服务器, 主要代码如下:

```
wget -O - >> /data/dyyx/probeData/sptanzhen_probeData.txt_59_yyyymmdd.txt
```

```
http://xxx.xxx.xxx.xxx:6555/probe/sp/59/xuanli/behavior/yyymmdd/probeData.txt.
```

创建表、FTP 采集、HDFS 加载以及执行 SQL 节点是对取过来的探针数据创建表结构, 然后一对一字段映射导入 HDFS 文件路径的过程。主要代码如下:

```
load data inpath '/user/jscn/dacp/sptanzhen/probeData/probeData.txt_*_yyymmdd'
```

```
overwrite into table jscn.STG_SPTANZHEN_LOG partition (dt=yyymmdd).
```

3 数据模型设计

数据模型主要包括业务生命周期基础数据模型、业务生命周期评价模型、业务贡献度评价模型。

3.1 业务生命周期基础数据模型

业务生命周期基础数据模型包括业务的覆盖用户数与终端数、订购用户数与终端数、点播终端数与次数、销账额以及基于以上指标求算的几个指标。以江苏有线某几个地市订购的某业务为例, 见表 2。

业务生命周期基础数据模型的数据是业务生命周期评价模型以及业务贡献度评价模型的基础。

3.2 业务生命周期评价模型

业务生命周期评价模型是对业务市场份额成长度、粘性成长度以及盈利分析能力成长度的分析评价模型。模型包括

点播量评价参数、点播用户数评价参数、业务销账评价参数、生命周期评价参数几个指标。

1. 点播量评价参数 (业务粘性成长度): $L1 = \text{评价当期业务月点播量} / \text{历史平均月点播量}$ 。

2. 点播用户数评价参数 (业务市场份额成长度): $L2 = \text{评价当期业务月点播客户数} / \text{历史平均月点播客户数}$ 。

3. 业务销账评价参数 (业务盈利能力成长度): $L3 = \text{评价当期业务月销账额} / \text{历史平均月销账额}$; 对于公益性业务, 可不考虑此参数。

4. 业务生命周期评价参数: $L = (L1 + L2 + L3) / 3$ 。

以某业务 4~7 月的业务生命周期基础数据为基础, 对其业务生命周期评价参数的几个指标量做了求算, 如表 3 所示。

根据以上数据, 可以看出这个业务处于较快的上升期, 点播量评价参数以及点播用户数评价参数逐月递增。需要注意的是, 这个业务的业务销账评价参数几乎无变化, 由此我们可以推断出此业务的铺开可能是因为有较大的优惠活动, 由此拓展客户。

3.3 业务贡献度评价模型

业务贡献度评价模型是对上线的业务之间做一个综合评估打分的过程。

1. 业务点播量贡献度: $C1 = \text{评价当期业务月点播量} / \text{平台增值业务月点总播量}$ 。

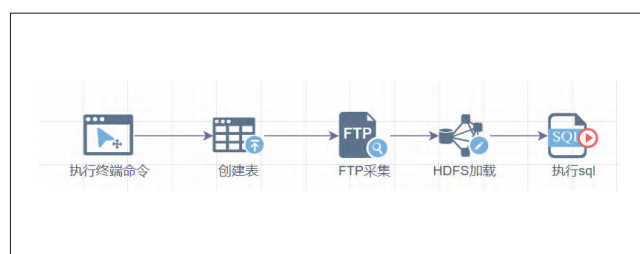


图2 探针ETL流程节点

表2 业务生命周期基础数据模型

业务名称	月份	覆盖用户数	覆盖终端数	订购用户数	订购终端数	点播用户数	点播终端数	点播次数	业务到达率	业务渗透率	业务活跃率	业务月销账额 (元)
业务	202204	3009363	3586342	204837	205237	15498	15926	139946	16.49%	5.72%	7.76%	3492983.05
	202205	2985745	3559066	211318	211769	20663	21237	198652	16.45%	5.95%	10.03%	3936785.80
	202206	2973514	3546313	218390	218888	28240	29233	2213019	16.43%	6.17%	13.36%	3824214.06
	202207	2960493	3531659	228682	229235	38946	40567	6011486	16.43%	6.49%	17.70%	4068237.60

1. 业务渗透率=业务订购终端数/业务覆盖终端数

2. 业务订购终端数=业务套餐订购终端数

3. 业务活跃率=业务点播终端数/业务订购终端数

4. 业务到达率=覆盖缴费终端数/全部缴费终端数

表3 业务生命周期评价模型

月份	业务名称	点播量评价参数	点播用户数评价参数	业务销账评价参数	生命周期评价参数	生命周期评价参数历史月均值
202204	业务 1	1.00	1.00	1.00	1.00	1.00
202205	业务 1	1.17	1.14	1.06	1.13	1.06
202206	业务 1	2.60	1.32	1.02	1.65	1.26
202207	业务 1	2.81	1.51	1.06	1.79	1.39

表4 业务贡献度评价模型

业务名称	业务点播量贡献度	业务客户数贡献度	业务收入贡献度	业务贡献度评价参数
电竞视频	0.02	0.03	0.02	0.02
名师空中课堂	0.03	0.09	0.03	0.05
上海文广	0.09	0.21	0.12	0.14
电影院线	0.02	0.21	0.13	0.12
芒果TV	0.38	0.17	0.23	0.26
学霸	0.02	0.02	0.04	0.02
炫力	0.43	0.17	0.43	0.34

2. 业务客户数贡献度: $C2 = \text{评价当期业务月点播客户数} / \text{平台增值业务月点播客户数}$ 。

3. 业务收入贡献度: $C3 = \text{评价当期业务月销账额} (\text{含套餐拆账销账额}) / \text{增值业务月销账总额}$; 对于公益性业务, 可不考虑此参数。

4. 业务贡献度评价参数: $C = (C1 + C2 + C3) / 3$ 。

基于业务贡献度评价模型, 我们对电竞视频等 7 个业务做了指标求算, 如表 4 所示。

根据以上数据, 可以看出电影院线这个业务虽然点播量占比不高, 但是对客户的粘性非常高, 且业务收入贡献度也较高, 需要继续大力发展。炫力业务无论是点播量、贡献量还是收入贡献度, 都占比非常高, 在增值业务中占据重要地位。

3.4 模型应用

对于业务生命周期模型以及业务贡献度评价模型这两个模块, 通过对其上线几个月以来存量业务的综合评估, 给业务部门提供了极具参考性的评估方案, 贡献度大或者用户粘合度大的业务得到了全力推广, 以这些业务为模板, 积极拓展新的业务; 同时针对部分长期持续贡献度很小的业务, 予以下线或者改版升级、内容迭代, 提升用户的体验度。

4 结束语

通过建立业务生命周期评价模型、业务贡献度评价模型, 为在江苏有线落地的增值业务提供了效果评估方法, 从而及时发现自身业务、合作伙伴业务之间的资源差异, 为江苏有线整体影视剧版权购买策略的制定提供重要的决策依据, 引导购买更符合大众需求的影视剧产品, 同时也激励内容服务提供商提供更高质量的影视剧内容, 合理规划新业务开展, 提高公司的运营成本投入和产出效益比。RTBE

参考文献:

[1] 孙婕. 一种针对电信产品的生命周期评估方法[J]. 中国科技论文在线, 2018: 1-8.
[2] 秦东旭, 徐瑾, 吕明, 张捷. 基于Hadoop的用户行为数据分析系统的设计[J]. 工业控制计算机, 2019, 32(10): 137-138.
[3] Q/JSCN 技 05.02-2020. 江苏省广电有线信息网络股份有限公司收视行为大数据采集接口规范[S]. 江苏: 江苏省广电有线信息网络股份有限公司企业标准, 2020.

第一作者简介:

王东明 (1985—), 男, 中级工程师, 硕士。主要从事广电行为数据采集、数据质量管理方面的研究。