



星环助力华数传媒大数据升级

近期,华数重磅推出阿拉丁推荐、数据罗盘、魔数水晶球三个主要大数据产品。

收视罗盘通过采集华数数字电视业务支撑平台的实时业务数据,经大数据平台一系列的复杂算法,运用实时计算技术,将业务运营的综合情况高效地展现出来。收视罗盘最大的特点是数据真实、实时高效。包括网内最热播的电影、电视剧、少儿影视、视频新闻、视频栏目、直播等业务的实时榜单和黄金时段、天、周、月等关键时间周期榜单,能在第一时间了解网内内容运营的总体情况。例如:在5分钟内,收视罗盘就能将公司各个业务平台上的运营综合情况高效地展现出来,使运营人员能第一时间了解全省内容运营的总体情况。

魔数水晶球通过大数据平台实时计算能力,运用大数据挖掘分析技术对华数门户平台上的用户行为进行综合分析,并做可视化展现,对用户行为进行归类 and 汇总,直观展示门户上用户的流向和对各栏目、页面的浏览情况,为运营、策划、编排人员提供从实时监测到定期分析的各维度业务指标。

自2014年3月将“大数据”首次写入《政府工作报告》以来,李克强总理在多个场合提及这一热词。大数据的应用在新媒体的运营中显得尤为如火如荼。优酷,腾讯,爱奇艺等互联网新媒体已经从技术层面上爬到了制高点,他们从直播,点播,轮播,时移,录制等基础功能结合用户评论,打分,投票等互动手段形成了大运营的概念。



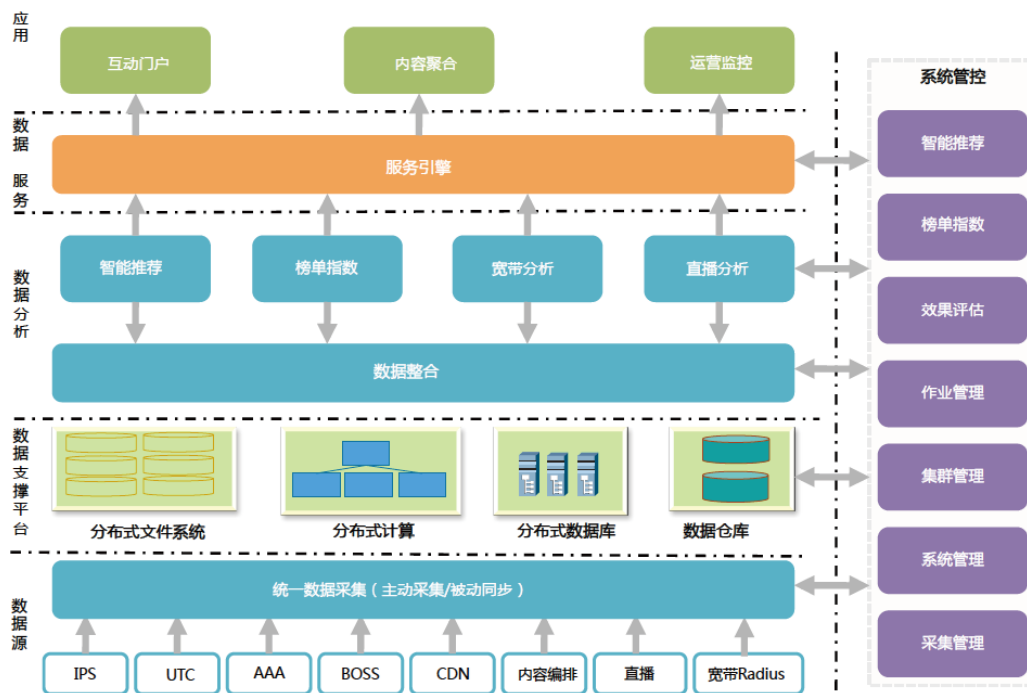
以优酷视频网为例，优酷的视频推荐中涉及上百个参数，每次要调整参数，都要手动调整十几甚至几十个参数，每天推荐视频的数据模型中要涉及的数据高达几十亿，一次小小的参数调整，带来的后果，可能是当天视频观看带来几百万的增长。美剧《纸牌屋》是其大数据分析的第一次战略应用。整部剧集是Netflix一次性在网站发布，供订阅者观看，完全颠覆了传统的剧集发布(每周一集)的模式。分析结果表明：热播三要素——凯文·史派西；大卫·芬奇和BBC出品三者结合，就可以吸引大量用户。由此可见，收视行为模式变更——越来越多的人不再像30年前那样，在固定晚上的固定时刻守在电视机前，等着收看电视剧的最新剧集，而是“攒”起来选一个方便的时间段在方便的设备上看。

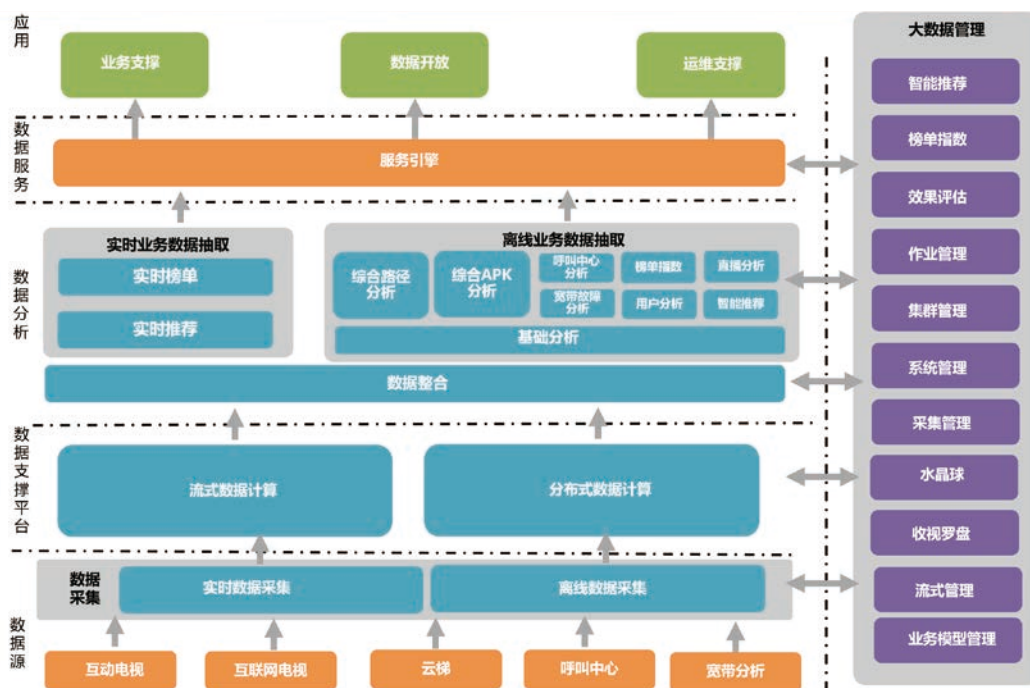
总的来说，新媒体行业由于用户粘性差，忠诚度低，所以各媒体除了拼上游内容资源外，更是不计成本的利用大数据的采集，计算，分析为大运营中的各类业务提供有效支撑，大数据技术的发展使得互联网媒体运营进入媒体运营的高级阶段。大数据造就了媒体运营的高级阶段。

早在2013年，华数传媒就依托星环科技的TDH Hadoop大数据平台完成了大数据从无到有的基础建设，完成了与门户，CLPS等系统的数据对接，实现了用户榜单，用户推荐等基本应用。可以说，华数传媒在广电行业的大数据领域一直保持着领头羊的地位。

然而，整个广电领域在技术与网络差异性越来越小的情况下，服务体验才是用户选择运营商的重要因素，而互联网新媒体的强力冲击使得广电行业必须拉近技术差距，华数作为广电行业的领头羊必须在用户体验服务上保持充分的竞争力，所以必须深挖用户价值，分析用户喜好，在现有大数据平台基础上做大做强，保持技术领先优势，保障市场竞争优势。

2014年，华数传媒大数据平台成功应对了服务性能增强、数据分析挖掘时间缩短、大数据业务支撑功能扩展与增加对网络、服务质量数据分析的四大挑战。两年时间过去，在发展日新月异的大数据领域，华数传媒又将面临新的市场、新的技术与新的挑战。基于此，有了华数传媒大数据三期的规划。





从华数现状分析可知，华数日志采集已基本完成互动电视域业务数据接入，互联网电视接入及部分运维数据（呼叫中心与宽带中心），但日志采集颗粒度较粗，无法为更精细化的数据分析提供支撑。所以在华数传媒的大数据三期规划中所面临的第一个就是数据采集与分析强化。

现有日志仅仅记录了用户观看点播，直播等业务的基础日志，并不能真正反映用户的真实喜好，应当对用户行为进行更精准的分析。更精细化的日志采集分析包括：用户浏览版位点击分布，用户页面停留市场，退出点，用户播放时的快进、快退、退出操作。

数据采集分析方面，用户主动发起的行为往往会比被动收集的信息更有价值。用户主动行为日志分析包括：户主动发起的评论，打分等互动行为，户通过呼叫中心报障的内容，联网用户的内容评论，打分数据融合。

此外，数据采集平台化的建设也是急需实现的目标。海量数据的采集与云化托管、服务已经成为一种热门业务，如AMAZON的REDSHIFT，华数在自身大数据平台基础上可考虑将现有采集能力平台化，提供数据采集，数据服务。

同时，功能化系统向运营化平台的转变也是必不可少的。可以看到，在华数传媒大数据平台的三个版本中，V1.0版本实现了华数大数据从无到有的基础平台建设，实现了基础的推荐，指数，榜单业务，完成了与部分业务系统的对接。V2.0版本在二期基础上实现了功能上的强化，对原有基础的推荐，指数，榜单业务分别进行了增强，实现了水晶球与收视罗盘的分析结果展现工具，实现了与部分运维数据的对接，实现了流式的实时计算分析。

而现有的平台则有四方面的问题需要解决与优化。

首先是集成度不够——目前仅实现了功能化，根据需求简单的将各功能堆砌成为一个系统，缺乏高可控的中心化平台以及松耦合的能力化组件。

其次是综合分析缺乏——先阶段各业务数据离散的为各类业务服务，分析的完整度和真实度不够，缺乏综合的全业务



数据分析。

另外，开放性——无论是市场需求还是政策发展需要，开放与共享都是平台化的重要特征，目前的大数据系统无法实现能力开放，外面的数据无法引入进来，内部的数据无法共享出去。

最后，扩展门槛高——由于组件能力化不够，无法用很低的代价实现横向的业务扩展，业务延展性不够，无法支撑华数大运营生态圈的发展。

所以，V3.0需要在原有功能化系统的基础上向运营化的平台转换，实现运算中心平台与采集、服务等能力化组件结构；实现横向业务扩展，包括管理端业务角色分权以及数据端分域；实现标准化的业务数据模型及对外服务数据接口；实现大数据综合数据统一分析。

大数据三期面临的最后一个挑战是需要更丰富的业务类型。现有大数据系统实现了推荐、指数、榜单等基础业务以及呼叫中心与宽带故障的运维运维，从一定程度上提高了转化率，节省了部分运维人力成本，但缺乏直观的运营指导，缺乏能够直接带来用户价值转化的业务，降低运维成本的相关业务深入不够。

其中，运营指导业务扩展包括过大数据综合分析为编排运营带来直观的数据指导，包括内容点击量，互联网相关内容热度，基于地域的内容热度，用户页面版位热度，用户页面浏览轨迹，用户观影行为统计等，通过分时，分段的对营销专题、活动等行为进行统计，形成周期性营销的有效运营指导，统计分析包括内容热度，页面喜好度等。

用户价值转化业务扩展包括对用户从开机到关机的整体行为轨迹追踪，对用户各环节的操作行为进行细化综合分析，刻画更完整的用户画像。对现有内容标签体系上进一步挖掘，结合互联网新媒体标签，将内容标签变为真正的可运营。通过用户完整画像与内容标签的结合，推出个性化广告，个性化推荐，订阅推广等新业务实现用户价值的转化。

运维成本降低业务扩展则包括了丰富现有运维数据，包括管理员操作分析，服务器报障分析，网络报障分析，数据库报障分析等多个维度。上述各类运维业务数据都汇总在大数据中心，经过综合统计分析后可提供运维分类的全局视图，有利于定位频发故障类型，频发地区，从根源上解决问题降低后期重复维护成本。

总得来说，大数据三期平台的建设目标是扩大数据采集范围，实现运营化平台，丰富新业务种类。

对于包括广电运营商在内的众多企业而言，精细化的用户需求掌控是未来发展不可变更的趋势。在此，Hadoop大数据平台的优势是不可替代的。华数传媒与星环,思华科技合力打造的大数据平台在过去的两年里经受了时间的考验，华数传媒也始终处于行业领头羊的地位。如今的大数据平台第三期规划将再次证明大数据时代的浪潮仍未过去。

星环信息科技（上海）有限公司

📍 地址：上海市徐汇区桂平路481号18幢3层301室（漕河泾新兴技术开发区）

✉ 邮编：200233 ☎ 电话：4008 079 976

🌐 网址：www.transwarp.io

TRANSWARP
DATA HUB