

ALIBABA CLOUD

阿里云

阿里云案例库
文化产业

文档版本：20211008

 阿里云

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
 危险	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 危险 重置操作将丢失用户配置数据。
 警告	该类警示信息可能会导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告 重启操作将导致业务中断，恢复业务时间约十分钟。
 注意	用于警示信息、补充说明等，是用户必须了解的内容。	 注意 权重设置为0，该服务器不会再接受新请求。
 说明	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击设置> 网络> 设置网络类型。
粗体	表示按键、菜单、页面名称等UI元素。	在结果确认页面，单击确定。
Courier字体	命令或代码。	执行 <code>cd /d C:/window</code> 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[] 或者 [a b]	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
{ } 或者 {a b}	表示必选项，至多选择一个。	<code>switch {active stand}</code>

目录

1.东方明珠：探索中国广电“智慧城市”未来转型路径	05
2.助力央视CCTV5直播2020年洛桑冬季青年奥运会	07
3.新华智云：基于实时计算打造舆情分析平台	09
4.NoxBi：基于MaxCompute构建全球化精准营销系统	14
5.新华智云：基于MaxCompute的媒体大数据开放平台建设	24
6.MORE VFX：阿里云桌面助力《流浪地球》视效艺术家“云复工”	27
7.品推宝：营销内容与流程管理平台上云	30
8.南瓜电影：实现移动端与服务器双向互通	33

1.东方明珠：探索中国广电“智慧城市”未来转型路径

东方明珠是上海的传媒旗舰，在文娱领域拓展的同时，也被广电总局赋予重任，早几年已经开始依托广电频谱开展下一代宽带无线网络试点，当时的业务主要侧重政府事业单位服务，基于业务的需求变化，东方明珠在宽带无线网络的基础上打造了一张全新的物联网网络及平台系统，成功实现了智慧城市商业模式和物联网行业应用的突破，非常值得同行借鉴学习。

作为国内唯一拥有完整文化产业布局的国有文化企业，东方明珠新媒体股份有限公司（SH.600637）旗下拥有媒体渠道、影视制作发行、版权供应分发、游戏、视频购物、旅游地产等丰富的业务板块。公司重点推进“智慧广电文娱+”建设，积极布局智慧城市物联网业务，从广电频谱这一老牌优势领域出发，不断寻求产业突破。目前东方明珠智慧城市采用最新技术搭建宽窄带物联网，业务已经覆盖上海市三个中心城区及临港、外高桥2个智慧园区，涉及超过46类应用场景。

发展阶段：智慧城市大脑的1、2、3、4、X成长逻辑

东方明珠智慧城市建设经历了几个发展阶段：2015依托广电频谱红利，建成了全国首个运营级的城域物联专网；2016年起在浦东新区和杨浦区进行双向频谱无线宽带网络建设，并以整体解决方案打造了物联网+社区精细化治理示范街道，2017年网络覆盖扩大到虹口区；2018年在普陀区建成上海首个区级“城市大脑”。截止目前，东方明珠已经建设538个基站，宽窄带两张无线专网覆盖普陀、杨浦、虹口三区，覆盖率达95%以上，安装完成20万余个传感器。同时东方明珠携手阿里云，引入阿里云IoT物联网平台，为政府社区的各类应用提供生长土壤。

以普陀区“城市大脑”项目为例，已建设完成覆盖区级、街镇（委办）、片区的三级总分管理平台和覆盖前三级及延伸至居委、物业、社区力量等的四级应用，实现“1、2、3、4+X”体系（1个城市大脑、2张专网、3级平台、4级应用、海量传感器和配套的处置流程），为高度精细化的政府管理提供了有利支持。

东方明珠的智慧城市应用建设获得当地政府及市民的高度好评。2017年东方明珠“下一代宽带无线”获得广电总局颁发的“广播影视创新科技奖”。2019年第15届深圳文博会期间，东方明珠现场展示了智慧城市应用，作为其在“智慧广电”建设中取得的阶段性成果，获得各方肯定。

创新突破：携手阿里云，打造物联网城市大脑平台，落地“智慧社区”

由于阿里云IoT物联网平台具备强大的开放性，可以解决协议适配、海量连接、数据存储、设备管理、规则引擎、事件告警等物联网应用开发和对接共性问题，东方明珠携手阿里云，以物联网平台进行城市社区智慧服务延伸，目前服务于普陀区的智慧城市大脑已经发展46类应用场景，比较典型的主要有安全管理类、居家养老类和社会共治类三大类。

其中安全管理类主要包括社区的单元门管理（物联网及手机门禁）、社区停车位及车辆违停智能管理、社区民防车库自动管理、社区消费安全监控、社区垃圾堆放智能监控、社区电梯运行监控以及针对建设工地的扬尘及环保管理和普陀河水水质智能监控；目前东方明珠针对社区服务的智慧城市大脑智能管理体系已在普陀区服务于20多个委办、10个街镇、278个居委会，128万人口。

同时，居家养老类则针对老龄住户的需求进行响应，主要包括社区老龄住户管理、老龄住户关怀监测以及住户和医疗资源打通的方面，通过小规模试点，东方明珠的智慧城市养老应用获得了社区老龄住户的普遍好评，联合试点的长征镇荣获工信部、民政部、卫生委颁发的国家第二批智慧健康养老应用试点示范街镇。

此外，在社会共治方面，东方明珠开发出顺道APP社区居民共治系统，该系统衍生于“智联城市大脑”项目，借助居民主人翁精神，共创美丽家园。顺道APP目前覆盖的报警干预处置人员超过2000人，对普陀区全覆盖，其他区部分街镇覆盖。通过政府授权，老百姓主动参与实现社会和社区的共治和共管。

行业发展：广电网络的转型与服务政府及社区紧密相关

2018年11月22日，广电总局印发的《关于促进智慧广电发展的指导意见》（以下简称《指导意见》）正式发布，提出了智慧广电发展的总体目标。在“加快智慧广电生态体系建设”方面，《指导意见》明确提出“加快广播电视与物联网等新兴网络业态的集成创新、协同服务”。

东方明珠和阿里的合作可以作为剖析中国广电“智慧城市”未来转型路径的典型样本。携手阿里云物联网开放平台，东方明珠将集聚更多物联网应用合作伙伴，实现更多垂直行业的突破，扩大应用的普及和覆盖。同时结合5G网络的建设，将广域覆盖的重量级、大带宽物联网应用与社区轻量级、低成本物联网应用结合，两者共同在智慧城市大脑IOT平台上形成优势互补和良性循环，共同推动智慧城市和智慧社区的应用走向纵深。

相关产品

- 阿里云物联网平台

物联网平台（原物联网套件）提供安全可靠的设备连接通信能力帮助用户将海量设备数据采集上云，平台提供功能丰富的设备管理能力帮助用户远程维护设备，平台提供丰富的API以及与阿里云众多云产品打通的规则引擎，帮助用户将应用快速集成。

更多关于阿里云物联网平台介绍，参见[阿里云物联网平台详情页](#)。

- 物联网络管理平台（Link WAN）

物联网络管理平台是组建网络的管理平台，符合LoRa WAN协议，支持网关管理、凭证管理、终端管理、数据安全等。用户可使用平台维护网络组网，也能与阿里云物联网平台搭配，组合成解决方案。

更多关于物联网络管理平台介绍，参见[物联网络管理平台（Link WAN）详情页](#)。

- ET城市大脑

ET城市大脑利用实时全量的城市数据资源全局优化城市公共资源，即时修正城市运行缺陷，实现城市治理模式、服务模式和产业发展的三重突破。

更多关于ET城市大脑的介绍，参见[ET城市大脑产品详情页](#)。

2.助力央视CCTV5直播2020年洛桑冬季青年奥运会

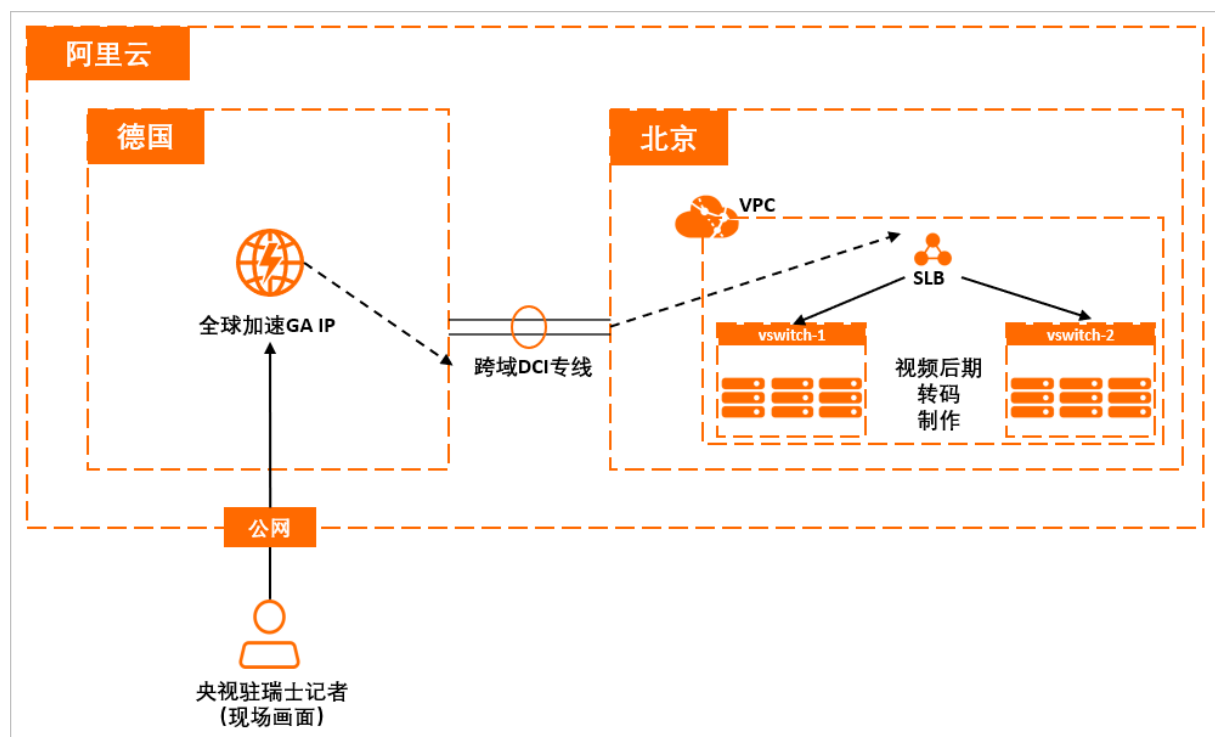
案例背景

2020年冬季青年奥林匹克运动会（2020 Winter Youth Olympics）于欧洲时间1月9日至22日（北京时间1月10日至1月23日）在瑞士洛桑举行。中央电视台体育频道CCTV5 进行全程直播通过阿里云网络全球加速（Global Accelerate）产品为直播网络保驾护航。

业务痛点

- 视频流畅度差：洛桑冬季青年奥运会直播采用4K高清视频，公网传输不稳定，会出现频繁卡顿、掉线现象。
- 部署成本高：传统卫星回传方案价格昂贵。
- 部署周期长：奥运会开展时间为14天，需要快速部署和扩容的能力才能保障赛事顺利进行。

解决方案



我们将采用GA将奥运会现场直播信号通过阿里巴巴优质全球网络传输回国，同时进行演播室电视直播，和CCTV5客户端互联网直播。

上云价值

- 高质量：
支持4K高清直播，大幅降低网络时延、减少网络抖动、丢包等问题。
- 低成本：
相比传统卫星回传方案，使用全球加速GA成本大幅降低。

- 易部署：

快速开通，无需复杂的部署过程，满足奥运会高效率的开通需求。

相关产品

- 全球加速

全球加速是一款网络加速产品，依托阿里巴巴优质 BGP 带宽和全球传输网络，帮助网络服务实现全球范围内用户的就近接入，减少延迟、抖动、丢包等网络问题对服务质量的影响，提升服务的用户访问体验。

更多关于全球加速的介绍，参见[全球加速产品详情页](#)。

公司介绍

新华智云是一家致力于通过大数据技术驱动媒体变革的公司，数芯是新华智云推出的实时舆情分析平台，旨在满足用户一系列舆情分析需求。包括：对事件、新闻、媒体、人物、地域、机构、行业，甚至关键词、热门话题等的抓取、识别、聚合、热度分析以及可视化展示等。

总览



事件跟踪



业务痛点

系统设计网络舆情是社会舆论的一种表现形式，通俗的说是人们通过网络表达对某些社会事件的看法和态度。网络舆情以事件为载体，以事件为核心，是广大网民情感、态度、观点的表达、传播、互动以及后续影响力的总和。一个舆情分析系统，主要解决的问题包括：发现事件、跟踪事件、发掘观点、评估影响力等。

- 舆情分析系统需要对接多个上下游。
 - 数据上游：爬虫数据和采购数据。爬虫采集不同数据结构类型的平台数据。
 - 数据下游：写入不同的存储系统。如统计结果写入RDS，清洗后数据写到表格存储中，文章存储到ES中做索引等等。

如果设计不合理，多上下游将会极大影响系统复杂度。

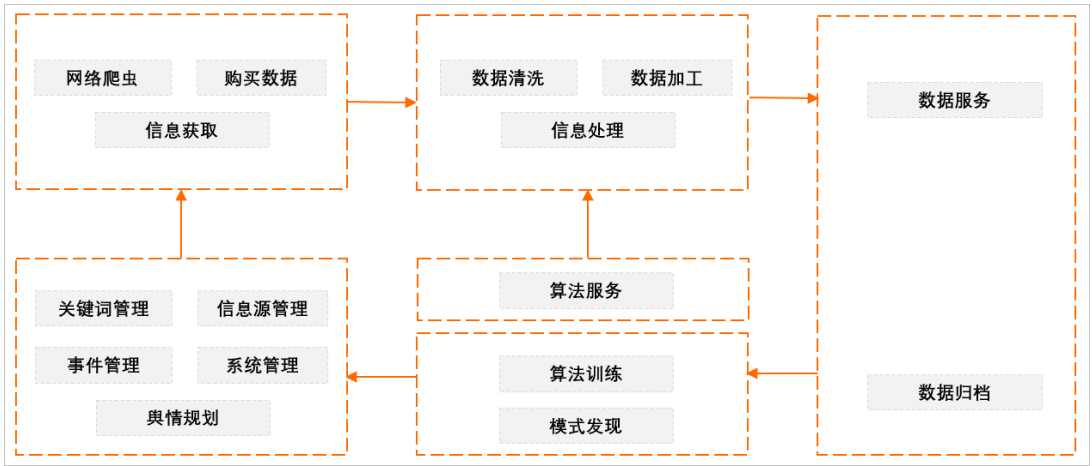
- 数据处理过程包括数据清洗和数据统计
 - 清洗过程既有基于规则的信息抽取，又有基于算法的实体识别。
 - 统计过程需要先把数据结构化拼装，然后按照各种维度进行灵活的统计。
- 具有自我迭代的能力

设计良好的舆情分析系统应该具有自我迭代的能力，能够根据历史数据进行优化，不断提供系统效能。

系统架构

大数据舆情分析系统架构图如下。

大数据舆情分析系统架构图

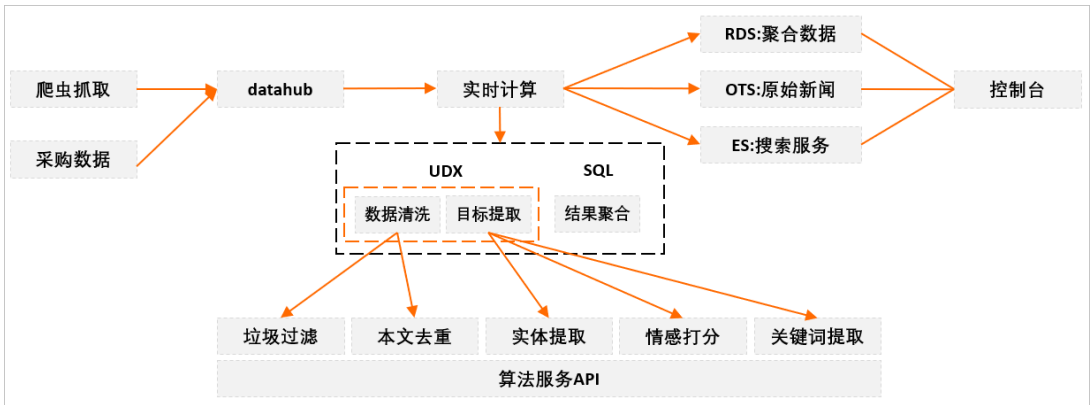


- 舆情规划是整个系统的控制器，控制着舆情发现的目标、信息源等。
- 信息获取是舆情分析的基础，能够快速、准确的获取足量信息是系统成功的前提。
- 信息处理与算法部分是舆情分析的关键，担负着把原始数据加工成信息与知识的重任，并且对归档后的数据进一步分析能够发现系统问题、发现新的事件和新的模型，能够进一步提升系统效果。
- 最终的信息要服务的形式的暴露出来，为人所用。

信息处理

数芯平台的信息处理部分如下图所示。

信息处理



- 数据来源（注：数据采集成功后统一发布到DataHub中）
 - 爬虫：实时的去各大网站、微博、微信等内容平台抓取数据。
 - 从其他机构采集数据。
- 数据处理
 - 实时计算订阅DataHub然后进行数据处理。数据处理包括两个重要的部分： 实时数据流和算法服务API。
- 实时数据流
 - 通过实时计算串起整个实时数据流，总体包括对数据进行清洗、提取目标，整理成结构化的数据，然后按照所需维度对结构化的数据进行聚合。并且要把原始数据、结构化的数据、汇总的结果都发布到下游存储系统中。
 - 数据清洗：对爬虫抓取的数据进行清洗，比如垃圾过滤、文本去重等。这一步通过实时计算提供的UDX（自定义函数）来调用算法服务API来完成。

- 目标提取：从已经完成清洗的数据中抽取出感兴趣的目标，比如抽取实体（人物、地点、事件等），对内容的情感评分，抽取文章关键词等，这一步的目标是从非结构化的文本中抽取结构化的信息。
- 结果聚合：对已经结构化的数据按维度聚合，比如某事件的按天计数，某新闻的传播热度等。最终把这些聚合信息联合原始信息分别写到下游存储中去。

- 算法服务 API

预先训练好模型，把垃圾过滤、文本去重、实体提取、情感打分、关键词提取等暴露成API服务，供实时计算调用。

- 数据存储与服务

RDS 保存聚合数据，表格存储保存原始新闻，ES对新闻建立索引，提供搜索服务，这三个数据存储直接为最终产品服务。

上云价值

数芯之前数据处理部分使用的是自建的spark，需要自行运维和对接各种上下游系统，迁移到了阿里云实时计算平台，整体收益包括：

- 运维成本：免运维，阿里云提供高保障。
- 对接上下游：直接注册，免开发。
- 开发成本：SQL开发，效率高，门槛低。
- 数据流：一个产品串起整个数据流，ETL用UDX，统计用SQL。

从更高的维度上看，这个案例属于实时ETL场景，实时ETL的目标把数据从a投递到b，中间进行清洗、格式转化、信息抽取等。如果对吞吐、实时性有一定要求，可以在方案阶段优先考虑实时计算产品。

相关产品

- 大数据计算服务·MaxCompute

MaxCompute（原ODPS）是一项大数据计算服务，它能提供快速、完全托管的PB级数据仓库解决方案，使您可以经济并高效的分析处理海量数据。

更多关于阿里云MaxCompute的介绍，参见[MaxCompute产品详情页](#)。

- 云数据库RDS MySQL版

MySQL 是全球最受欢迎的开源数据库之一，作为开源软件组合 LAMP（Linux + Apache + MySQL + Perl/PHP/Python）中的重要一环，广泛应用于各类应用场景。

更多关于云数据库RDS MySQL版的介绍，参见[云数据库RDS MySQL版产品详情页](#)。

- 表格存储

表格存储（Tablestore）是阿里云自研的面向海量结构化数据存储的Serverless NoSQL多模型数据库，被广泛用于社交、物联网、人工智能、元数据和大数据等业务场景。提供兼容HBase的WideColumn模型、消息模型Timeline以及时空模型Timestream，可提供PB级存储、千万TPS以及毫秒级延迟的服务能力。

更多关于表格存储的介绍，参见[表格存储产品详情页](#)。

- 阿里云·Elasticsearch

提供100%兼容开源Elasticsearch的功能，以及Security、Machine Learning、Graph、APM等商业功能，致力于数据分析、数据搜索等场景服务。

更多关于阿里云 Elasticsearch的介绍，参见[阿里云·Elasticsearch 产品详情页](#)。

- 实时计算

实时计算（Alibaba Cloud Realtime Compute, Powered by Ververica）是阿里云提供的基于 Apache Flink 构建的企业级大数据计算平台。在 PB 级别的数据集上可以支持亚秒级别的处理延时，赋能用户标准实时数据处理流程和行业解决方案；支持 Datastream API 作业开发，提供了批流统一的 Flink SQL，简化 BI 场景下的开发；可与用户已使用的大数据组件无缝对接，更多增值特性助力企业实时化转型。

更多关于实时计算的介绍，参见[实时计算产品详情页](#)。

4.NoxMobi：基于MaxCompute构建全球化精准营销系统

关于NoxMobi

NoxMobi是领先的数字营销和广告平台之一，通过我们强大而独特的技术驱动的人工智能（AI）和大数据（BigData）商业解决方案，为广告商和出版商提供服务。NoxMobi智能移动广告平台将AI+大数据+广告相结合，构建了一个具有不断增长的用户数据库的国际广告系统，为广告商提供高精度、高效率的目标流量。我们的全球交通覆盖了前五大地区，包括亚太地区、东南亚、欧洲、北美和南美洲。NoxMobi是为国内应用程序收购高质量海外用户的优质渠道。我们不仅为世界带来了优秀的产品，而且还带来了先进的技术和新的商业模式。我们的AI+大数据努力引领了应用市场营销的创新。

什么是数字营销

目前全球广告市场规模约6000亿，其中约30%为互联网和移动互联网广告，所以这块蛋糕也是很大的。数字营销、互联网广告、在线广告、计算广告、程序化广告等所说的基本上都是同一件事情，这些主要解决的问题就是如何在互联网媒体上投放广告的问题。

数字营销主要有三方参与者：

- 广告主（需求方）：其需求往往是面对流量和库存的，其最大需求就是如何成本更低，效果更好。而效果的定义往往是不同的，一般分为品牌和效果两种。对于品牌而言，比如可口可乐打广告，人们看到广告可能不会立即去买一瓶喝，但是品牌广告会深植在人们的心里，进而产生长期的效应；而效果广告比如在头条上看到一个广告，用户当时看到有兴趣可能就会点击之后进行下载了，这样就能发生直接的转化效果，这种就叫做效果广告。
- 媒体（供应方）：其需求比较简单，就是如何收益更高。收益具有长期和短期之分，媒体往往不能因为短期的高利益而放弃媒体的形象，这样才能让长期的利益更高。
- 中间商：主要就是通过连接供需来赚差价。有些广告宣传的中间商不赚差价基本上是不可能的，只不过是赚和什么时候赚的问题。赚取差价的主要手段就是规模效应和垄断，而无论是广告主和媒体只要能够垄断就能够具有更高的议价权。

总而言之，数字营销就是应用最新的互联网技术来提高营销效率，会应用比较前沿的人工智能、机器学习、经济学上的博弈论等。广告主会追求减少浪费，让最合适的人看到广告，这也是数字营销相比传统营销的特点。媒体则是将广告为卖给最需要的人，最需要的人则将会通过出价体现自己的需要。中间商则是高效地撮合交易。如下图所示的是广告行业的简单生态图，图中的DSP大概就是前面提到的需求方，也就是广告主。其右边是流量方也就是媒体，下面这部分则是赚取差价的中间商。



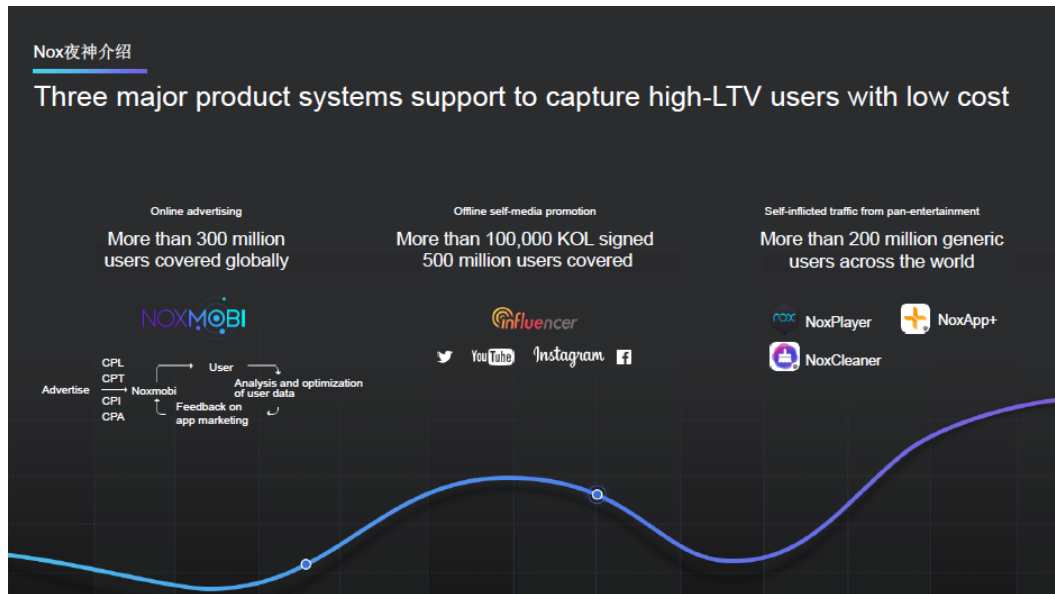
NoxMobi的广告系统

NoxMobi的广告系统包括夜神、NoxInfluencer等平台。

Nox夜神的主要优势：

- Nox 夜神：Three major product systems support to capture high -LTV users with low cost

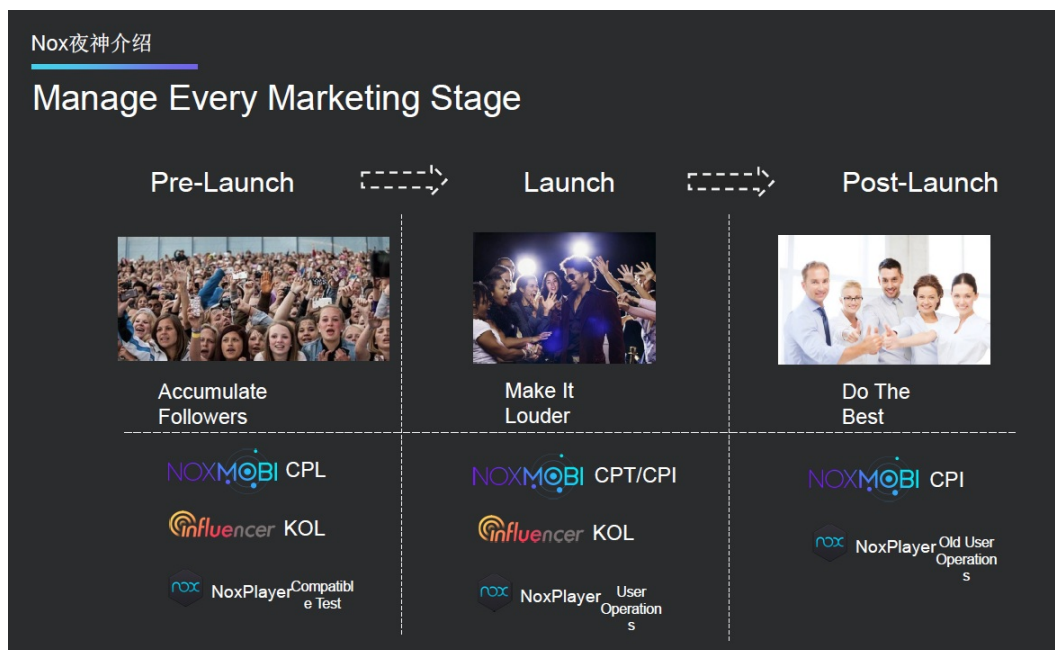
NoxMobi：广告和业务系统



上图中最左边是Noxmobi 广告平台；中间的是Influencer，在国外基本上就是“网红”平台；最右边的NoxPlayer 和NoxCleaner 就是Nox 自有的媒体流量，多点在线的广告也主要会投放在这些广告上。

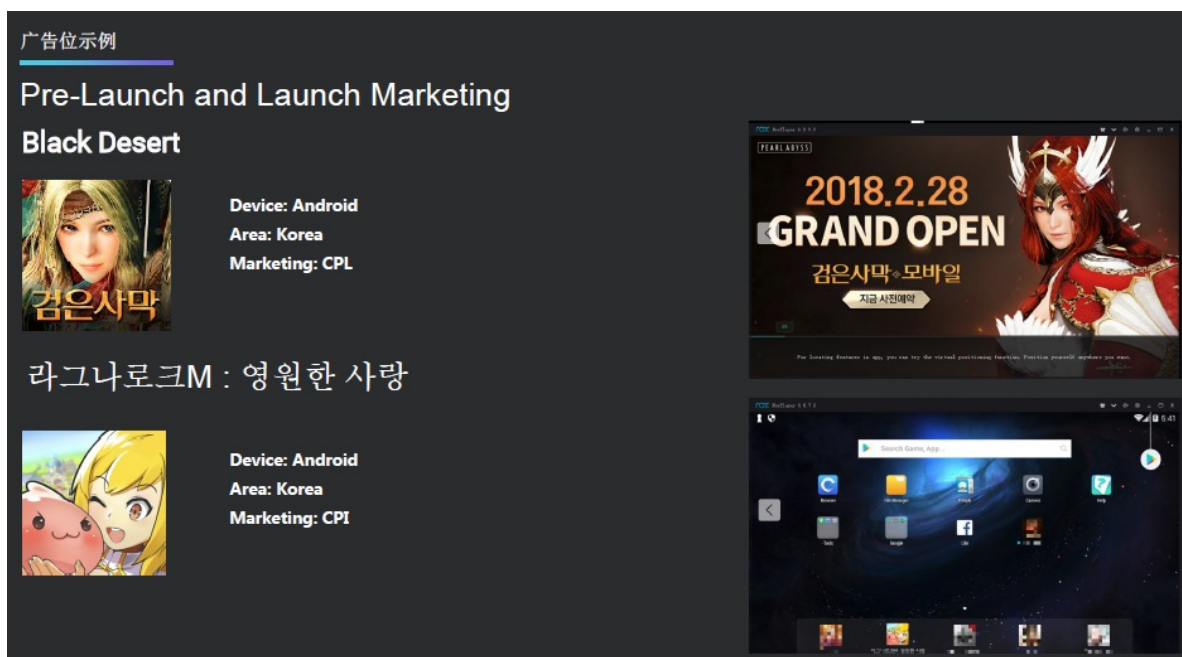
- Nox 夜神：Manage Every Marketing Stage

NoxMobi: 夜神2



如上图所示的是夜神在发行前期、中期和后期所应该使用的相应产品。

- Nox 夜神：Pre-Launch and Launch Marketing



如上图所示的就是Nox 流量上的广告位，图中右侧是模拟器的启动图和 Launcher 部分的广告位，这些广告也不会非常影响用户的体验。Nox 主要是做海外的生意，而国内和海外的广告市场是不同的，所以也是分开来做的。

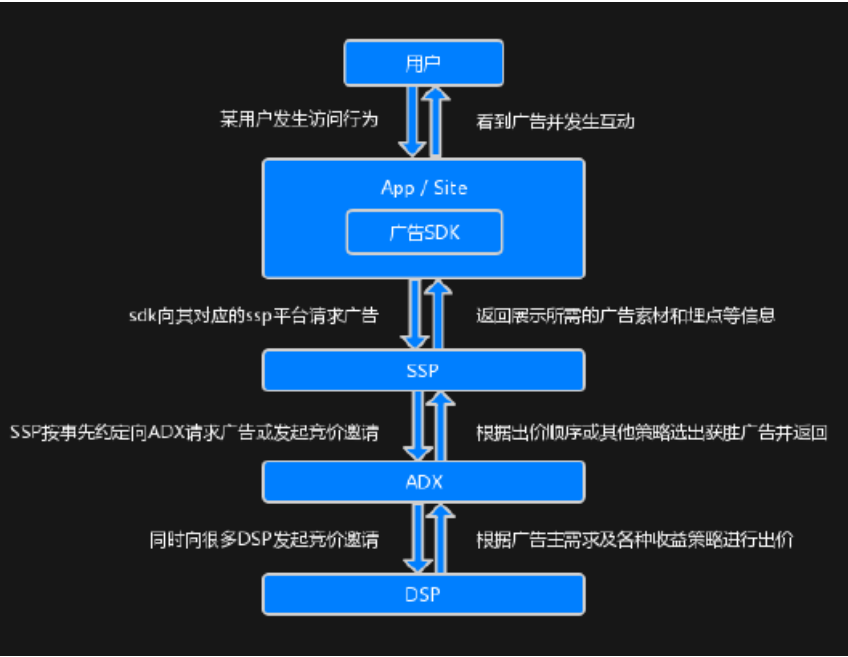
NoxInfluencer 示例



如上图所示的是目前主推的Influencer，目前也有一波流量红利，找网红推送 APP，右图是Nox 为Keep 做的广告，比如在健身教练所发出的视频或者直播，给教练一些收益，一起来做视频，也可以看到Keep 的广告收益也是非常好的。目前 Nox 也对于自己的流量端产品用Influencer 进行营销，花费大约10 万，使得自然新增达到了每天1 万，APP 开发者都知道这其实是一个非常可观的数字，因为自然新增能够将应用在GooglePlay 上的排名推到很高的位置，使得NoxCleaner在一些国家直接达到工具排行榜的前三。

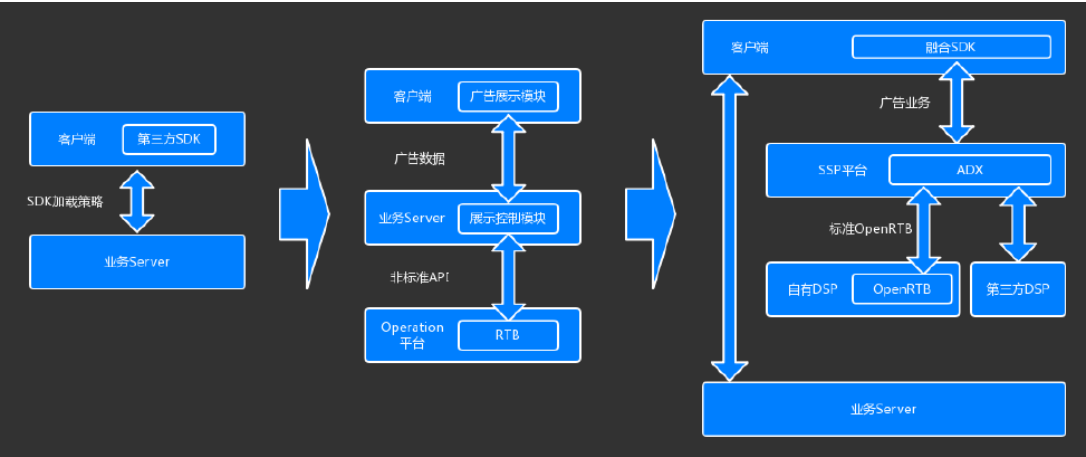
NoxMobi广告业务流程及演进

广告业务流程介绍：Online advertising



在线广告业务的基本流程可以从用户的访问开始，用户首先打开一个APP 或者 访问一个页面，当发生了用户访问行为之后，一般在APP 里面会有广告的广告SDK，之后SDK 就会触发展示然后去请求广告的内容。SDK 一般会与一个SSP 关联，属于 某一家SSP，然后去SSP 上请求广告。SSP 再对接ADX，也就是之前所提到的中间商，其主要负责撮合交易。ADX 又会对接很多的DSP，并向很多的DSP 发起竞价邀请，DSP 则将会代表广告主决定是否需要投广告以及广告的出价是多少，并在 ADX 进行多家的比较，做一次竞价的拍卖，价高者得，但是按照第二高价收费，再将获胜者一步步返回给用户，用户最终就会看到赢得本次竞价的广告。广告后续还会发生一些点击、下载、安装等行为，并进行进一步跟踪。

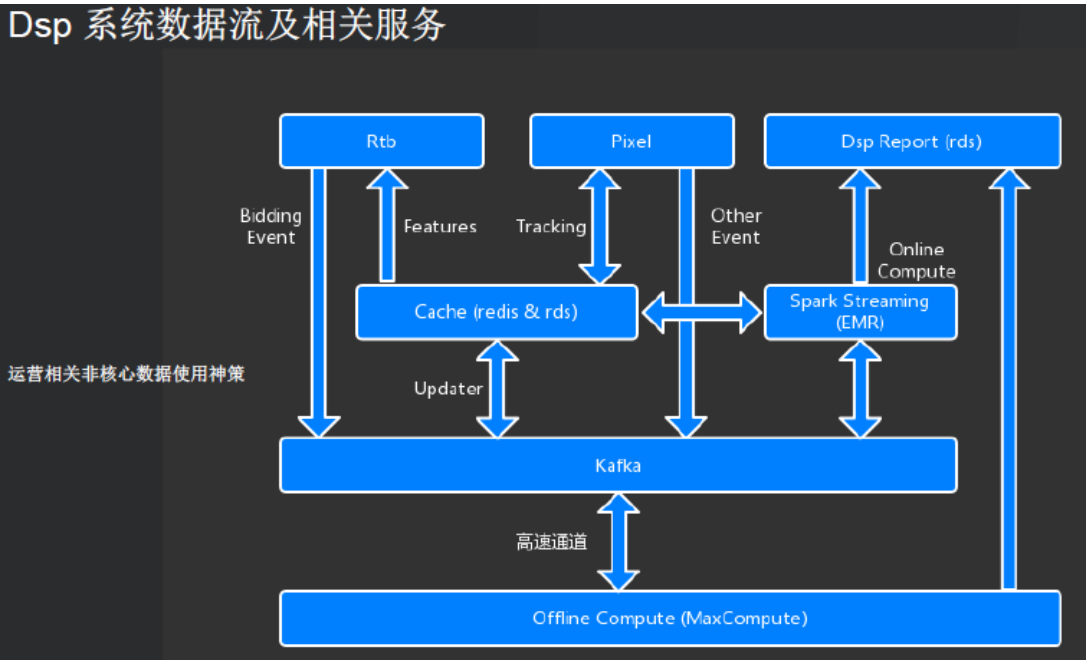
Nox广告系统演进



如上图所示的是Nox 广告系统的演进情况，首先可能与小开发者一样直接集成第三方的SDK，而第三方的SDK 的广告是完全不能被控制的，只不过会每天给 Nox 一定的收益，而在这个阶段其实收益也是比较低的。后来，Nox 就自己做了一个广告展示模块，由业务同学去接国外的离线Offer，在国外某些广告主或者品牌会放出来一些离线的Offer，这样的收益就比第一阶段高出了很多，这个阶段的系统会自己在客户端做展示模块，而在业务Server 里面则会有相应的展示控制模块，再往后就是在接到Offer 之后做一些简单的CTL 预估，之后再做一些比较并将收益比较高的投放出去。第三个阶段系统展现图中的中心部分就是广告业务，最外层的两个模块就是原来的业务系统，最外层两个模块之间的连接就会变得很细，基本上只做流量切分的工作，而中心部分就是前面所提到的DSP、SSP 以及集成的ADX 模块，从而形成一个完整的广告生态。在这里，除了自己本身的DSP 还会和第三方的DSP 进行竞价保证流量的收益最大化。此外，SDK 还可以做成融合的，虽然其不开放，

相关技术及MaxCompute 应用

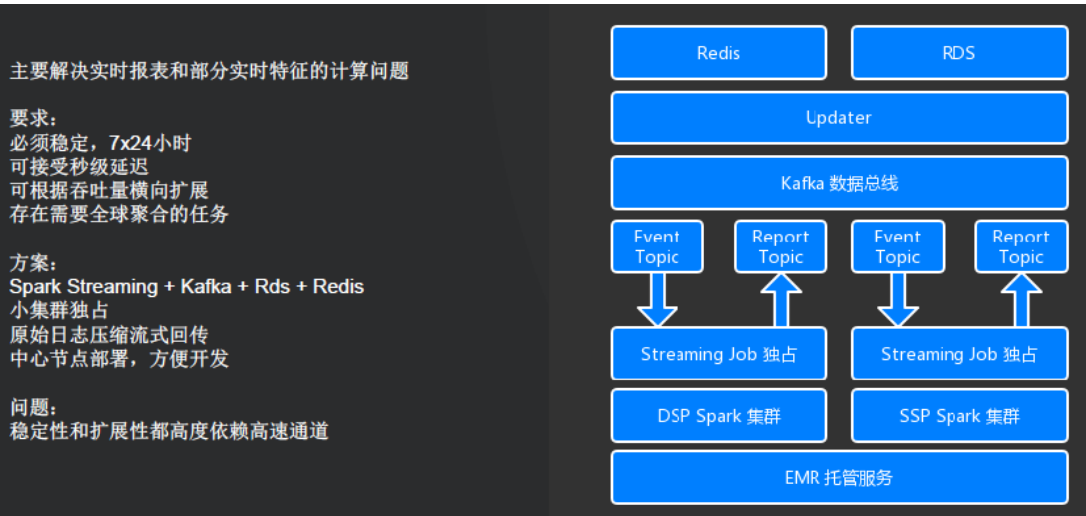
Dsp 系统数据流及相关服务



SSP 和ADX 相对而言比较简单，因此在这部分将会重点讲述一下DSP 系统数 据流。Rtb 就是Runtime Bidding，也是DSP 里面比较重要的一个部分，主要是做 实时竞价的。所以广告流程一般而言就是从Rtb 模 块开始，在竞价的过程中会产生 Bidding 的Event，并将Bidding 的Event 都输入到Kafka 里面，Kafka 里面 会有 订阅的消息，一直更新大的Cache，这里面的数据会通过流式计算做成feature 再反 馈给Rtb 系统。Pixel 就是事件服务，比如发生了展现或者安装之后就会访问到这个 服务，而这个服务将会Tracking 到本次 广告的所有Session，这样的数据也会流入 Kafka，同样由Updater 和流式计算进行处理。而从Kafka 里面会 另外分出来一只 数据流通过高速通道回流到中心节点，也就是MaxCompute 上面。MaxCompute 会进行一 些离线的报表计算和特征，报表就会输出到DSP Report 上面（比如 RDS）。

其实Nox 设计的所有广告的核心数据将会走这样的一套比较复杂的流程，而一 些运营相关的非核心数据则会 主要使用了自己搭建在阿里云上的神策这款BI 服务， 其也属于企业级的应用服务。

流式计算Spark Streaming应用

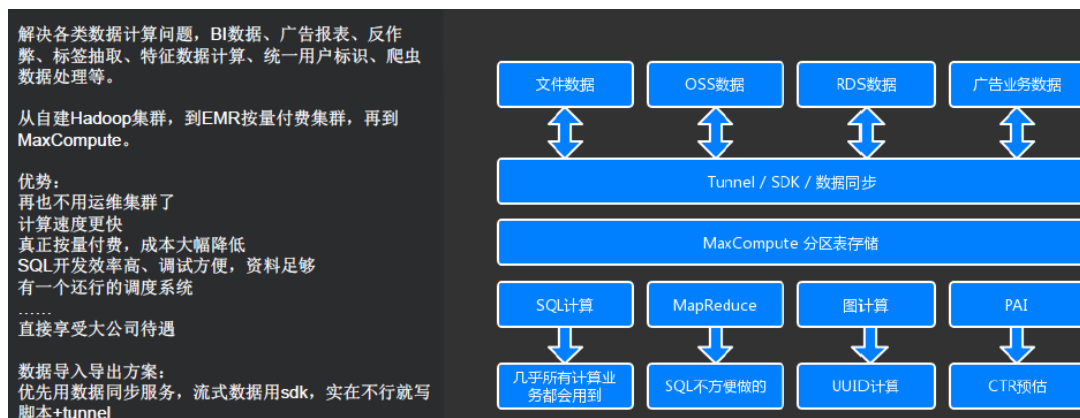


流式计算Spark Streaming 主要用于实现实时的报表以及实时特征的计算。因为业务的主要要求是必须稳定并且能够实现7*24小时的可用。可以接受秒级延迟，比如广告投出去了，晚10秒钟展现在报表里也是没问题的。可根据吞吐量横向扩展，比如突然新接了几家SSP，突然变得流量很大，不能在这个时候让系统挂掉。此外，因为业务在全球都有，所以需要全球的聚合任务，需要通过一个平台看到各个国家的数据。

Nox选择的方案就是：Spark Streaming 能够将上面几项需求全部满足，另外就是配合Kafka、RDS以及Redis做输出。在部署上面，需要实现小集群独占，这里所用到的就是阿里云EMR，其可以帮助客户托管集群，Nox只需要在阿里云EMR上面申请一个小集群，比如三到五台机器，这些机器申请之后就不再释放掉了，会一直独占着，并且7*24小时地跑流式计算任务。原始日志压缩流式回传，这个是因为Nox在各个数据中心都有Bidder或者Pixel的服务，会产生很多数据，之前的一种方案是在每个中心先将数据计算成半成品，之后在进行回传，这样所用的带宽就会比较小，但是如果采用这样方案，那么所有的功能都需要开发两套，在本地先计算，之后传回来再进行聚合计算，这样就会比较复杂，因此最终决定将日志进行压缩，以流式方式进行回传，这样的方案在验证之后发现所占的带宽不是很大，而因为是流式传输，因此带宽也比较平稳，虽然这里所用的带宽属于高速通道带宽，因此成本也可以接受。而压缩则使用了Kafka，其是能够支持压缩协议的。此外，中心节点部署能够方便开发。

上图中最底层就是阿里云的EMR托管服务，在其上是DSP平台和SSP平台，他们的集群是分开的，如果流量特别大，某一个平台被打挂掉了，另外一个平台是不会受到影响的。而托管服务的好处就是能够托管很小的集群，对于企业而言也没有什么成本。Kafka里面输入的就是Event的Topic，之后还会输出回Kafka，这样Updater再将Kafka里面的数据放到Redis或者RDS用于构建模型和计算报表。这样的设计，唯一的问题就是比较依赖于高速通道，这样稳定性和扩展性就有可能受到限制。

离线计算MaxCompute应用



离线计算部分，Nox主要使用了MaxCompute。几乎使用了MaxCompute来解决各类数据计算问题，BI数据、广告报表、反作弊、标签抽取、特征数据计算、统一用户标识、爬虫数据处理等。其实一开始，Nox也是自建Hadoop集群，购买了阿里云的ECS搭建集群，从最开始的6台一直到后来的十几台，这时候实在扛不住了，机器经常宕机，因为使用的是Spark，因此内存很容易占满，某一天用户突然增多了，数据就没了。此外，这样的成本也非常高，因为当时主要运行BI数据，所以基本上都是在晚上运行的，而白天机器则处于空闲状态，因此成本很高。后来采用了EMR的按量付费集群，晚上申请之后跑数据，但是白天能够释放掉，但是这样的过程则比较漫长的，需要10到20分钟。后来Nox开始接触到MaxCompute，使用起来非常好，其带来了许多优势。首先，不再需要运维集群了，此外其计算速度很快，虽然说Spark的计算速度很快，但是小集群的Spark和大集群的Hadoop是无法比拟的，所以大集群的Hadoop其实计算速度是很快。MaxCompute是真正的按量付费，因此成本也能够大大降低，而自建Hadoop、使用EMR以及使用MaxCompute的成本是成量级降低的。差距也是非常大的。主要使用SQL开发，效率比较高，也便于调试，文档也比较清晰。此外，MaxCompute还提供了一个还不错的调度系统，如果是自己搭建这样调度系统还是比较困难的。

对于数据的导入和导出而言，因为Nox有很多海外的服务，有些服务是不能覆盖到的。所以Nox采取的策略是优先使用数据同步服务，而流式数据则使用SDK，当数据同步和SDK都不合适就写脚本+tunnel导入和导出数据。

如上图所示的文件数据主要是爬虫，因为一些服务的日志是达到OSS上的，并且有一些外部数据也是先上传到OSS上面的。RDS则是什么都有的，广告业务数据就是前面所提到的，这些数据会选择合适的方式统一进入到MaxCompute的分区表里面。SQL计算基本上都会用到，MaxCompute则是在SQL写起来很费力或者运行很慢的情况下使用。图计算使用的并不多，只是会在计算同一用户UUID的情况下使用，这里应用了最小连通域的算法，比如一个用户使用了多个设备，则需要将这些设备统一地关联到同一个用户身上。而PAI平台对于广告的CTL预估非常重要。

特征计算和标签抽取

相关技术介绍

特征计算和标签抽取

大部分特征和标签使用SQL计算

定义特征，最近一周内活跃天数，则有0~7的取值

定义标签规则，例如：
一周内活跃 0天、1天、2~3天、4~5天、6~7天的分别是不活跃、低活跃、中活跃、高活跃、极高活跃用户

做好定义，然后就是展现SQL技巧的时候了
多用with，注意代码风格

优先使用内建函数，无法满足的时候考虑使用UDF、UDAF、UDTF，还不行或者跑的太慢的时候使用MapReduce，比如对一大批特征做等频离散化

尚未实现平台化，还需写代码，好像有个服务能实现上面的功能，待试用

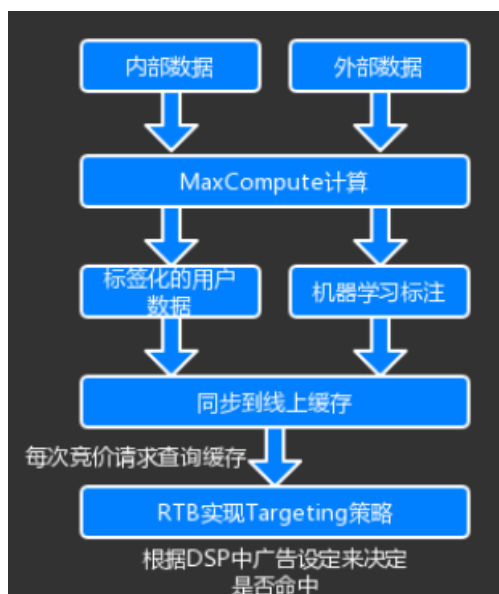
某第三方DMP对外标签体系（部分示例）

标签大类	标签	标签值（示例）
人口学	性别	男、女、未知
人口学	年龄	年龄段
人口学	消费能力	超高、高、中、低
人口学	作息时间	作/息
设备信息	激活时间	时间戳
设备信息	设备型号	设备品牌&型号
设备信息	操作系统	操作系统版本
设备信息	操作系统语言	语言编码
设备信息	真机概率	概率值（数值）
设备信息	羊毛党预警	概率值（数值）
设备信息	视频会员	是否
设备信息	所属运营商	运营商
设备信息	3/4G触点占比	占比
关联设备	家庭-OTT	OTT-ID
关联设备	移动设备	设备ID
关联设备	电脑	CookieID
关联设备	同住人	SuperID
地理信息	常住城市	城市
地理信息	工作区域	区域
地理信息	工作流动性	固定/同城流动/跨城商旅
地理信息	居住区域	小区
地理信息	休息时段活动区域	区域
地理信息	深夜活动	经常、偶尔、无

如下图所示的某第三方DMP的对外标签体系的示例，大概分了几类，比如人口学、设备信息等大类，在每个大类下面还会有多个标签。特征一般而言就是连续值，标签则是指将连续值做一些规则之后所打的标签。举例而言，定义特征，最近一周内活跃天数，则有0~7的取值。而定义标签规则，则是一周内活跃0天、1天、2~3天、4~5天、6~7天的分别是不活跃、低活跃、中活跃、高活跃、极高活跃用户。当在做好定义之后，就要看大家的SQL写的是好是坏了。之前在不支持with的时候，SQL代码一般都要写很大一堆，而且很难改动。此外，在写SQL的时候注意代码分隔还是很重要的。另外的一些建议就是优先使用内建函数，虽然一些内建函数和UDF的功能差不多，可能一个UDF能够实现两三个内建函数的功能，但是效率却相差了很多，虽然使用内建函数会让代码看起来丑一些，但是绝对比UDF运行速度要快得很多，所以在内建函数无法满足需求的时候再去考虑UDF，实在不行就可以用MapReduce实现，比如对一大批特征做等频离散化。

上述这些都是DMP的标准功能，但是Nox目前还没将其实现平台化，都是使用标签写的。而阿里云上有标签服务，目前也在考虑使用。

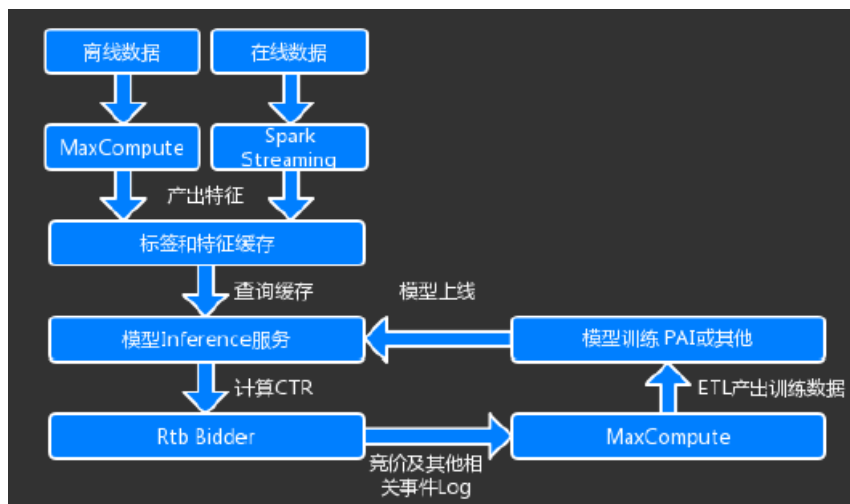
Targeting



所谓Targeting 就是人群定向，相对于把特征输入模型而言，标签式Targeting 主要是方便人来操作，使用投放人的经验通过标签定向的方式来优化表达。比如在 广告投放初期可以比较好地缩放盲打范围。另外一种则是Look-alike 方式，Lookalike 方式的定向为寻找相似的人，种子用户为正例，从所有用户中找到正例概率较大的人群。以上就是定向的两种主要方式。

如上图所示的定向的主要做法就是将内部和外部的数据输入到MaxCompute 里面，经过各种计算将标签化的数据或者用机器学习标注好的数据同步到线上并缓存 好。之后在进行实时竞价RTB 里面查询缓存，命中之后就在DSP 里面由广告主配置，命中了就投放广告，否则就不投放。

Ctr预估



Ctr 预估是Nox 投入比较多的一项工作。Ctr 预估并不一定是预估Ctr，还可 能去预估Cvr 甚至是Ctr 和Cvr 的乘积，这些统称为Ctr 预估。其作用是首先计算 Ecpm 的值，Ecpm 就是每次展现的期望收益，期望是一个概率论上的概念，其等于 用单价乘上本次收益可能的概率。Ecpm 也将指导绝大部分投放相关策略。

如上图所示，主要分为两条线，一条是离线数据会走MaxCompute，而在线则会走Spark Streaming。总体最后会输入到标签和特征的大Cache里面，Cache中的数据有一部分直接加载到内存里面，另外一部分比如用户特征无法加入内存就会在Inference服务查询缓存，这里就会组合出一个特征的向量，用来计算Ctr的值，并将值返回给Rtb的Bidder，在Bidder做一系列的策略，最终得出竞价的决策。竞价完成之后，是否参与、是否竞价成功以及是否展现等日志都会灌入回MaxCompute或者Spark Streaming进行模型训练，最终对已有模型进行更新，形成一个整体的闭环。Online model需要使用Spark Streaming，而Deep learning就需要用TensorFlow等了。

Pacing

Pacing比较复杂一些，其就是不止考虑单次展现的收益，而要在单次竞价时考虑对全局收益的影响，比如考虑一天之内总收益如何，比如可能将转化率最高的Offer在一天开始的两小时内都投完了，但是其他的Offer都没有投出去，这样计算下来总收益并不如将全部广告都投完的收益高。Pacing的整体思路就是通过流量分层和分时的统计和预估，用数学方法来保证全局收益的最大化。Nox则根据Yahoo的论文实现了自己的方案，这里面最核心的就是将Ctr估算准确，并将分层和分时的各种统计值计算好，然后按照其策略执行即可。

Nox目前也在寻求更多的合作伙伴，希望更多与具有出海意向的开发者进行深度合作。

5.新华智云：基于MaxCompute的媒体大数据开放平台建设

案例背景

随着自媒体的发展，传统媒体面临着巨大的压力和挑战，新华智云运用大数据和人工智能技术，致力于为媒体行业赋能。通过媒体大数据开放平台，将媒体行业全网数据汇总起来，借助平台数据处理能力和算法能力，将有价值数据内容和能力开放给用户。本文主要从新华智云数芯平台，媒体行业数据特征，批流处理数据架构，以及通用的媒体大数据平台能力等几个方面介绍了如何基于MaxCompute做媒体大数据开放平台建设。

关于新华智云

新华智云是新华社与阿里巴巴合资成立的一家公司，主要面向运用大数据和人工智能技术，致力于为媒体行业赋能。随着自媒体的发展，自媒体公司非常的火热，如今日头条，抖音等。对传统媒体来说，面临着巨大的压力和挑战。传统媒体急切的希望互联网的技术帮助他们赋能。

媒体大脑是新华智云底层的产品品牌，数芯是媒体大脑基础的数据平台。数芯定位为媒体大数据开放平台，平台中包含了媒体行业所依赖的各种各样的数据。数芯希望通过平台本身的数据处理能力和算法加工，将挖掘的有价值数据内容和能力开放给上层用户。目前，数芯包含了各种各样的数据，涵盖了400多万互联网站点，主要以中文站点为主，日增7千万文章，包含微信公众号，微博账号，图片以及多媒体来源等。将不同来源存储在平台中，之后对接数据，进行更多的加工和运用。“数据已经不再是一种成本，而是成为一种生产资料”，这句话在媒体行业更能够有所感受。一篇新闻除了是新闻之外，还是一条数据，同时新闻本身又会衍生产生新闻。很多媒体公司利用互联网技术获得大量的新闻内容之后，辅助他们新闻的生成过程。数芯会将媒体数据从不同的源头，不同供应商，以不同内容格式存储到平台中，进行数据清洗，结构化，加工等一系列的操作之后，生成有价值的内容。如，这段时间之内媒体关注的热点，互联网上发生的事件，同时识别新闻和机构之间的关系，基于实体将各个渠道上的数据整合起来，为用户提供统一的视图。用户可以看到信息的发酵过程，以及信息在什么地方传播。

数芯- 媒体大数据开放平台

数芯是一个开放的大数据平台，开放的点主要有三个方面。

- 第一，数据的开放。任何数据进到数芯之后，都会在整个大数据处理的链条中做计算，内容结构化，加入标签。同时基于用户感兴趣的数据范围，做标签的特征过滤，筛选用户想要的信息。数芯帮助用户了解互联网上的信息，了解互联网上与自身相关的事件。
- 第二，数芯提供智能能力的开放。用户不管怎么努力，不可能获取到互联网上所有的数据；而且不可能每个公司都会自建内容大数据平台，做数据分析，也不可能都配备算法工程师；而且公司的数据处理能力也不是足够强；媒体大数据平台可以帮助用户将与他们有关的数据加工好，通过算法能力获得用户关心的信息。如用户提供一个文章，数芯可以反馈这篇文章与谁相关，相同的文章在什么地方传播，文章由谁编写，发在什么地方等等信息。通过开放算法能力，帮助用户做数据能力和算法能力的应用，提供文本反垃圾服务，互联网内容结构化服务，文本内容实体识别服务，文本去重判定服务，图像人物识别服务和图像标签化服务等等。
- 第三，产品能力的开发，包含舆情监测能力，版权检测，传播分析等能力。
- 媒体数据特征
 - 媒体数据是非结构化性非常强。与传统行业数仓最大的区别是媒体行业90%以上数据都是非结构化的，如文本、图片、视频等等。
 - 数据来源多样性。媒体行业数据有多种外部的数据源，它们有不同的数据提供的方式，提供不同的数据能力。所以必须要有强大的数据汇总能力才能将所有数据汇总在一起并服务好下游客户。

- 数据时效性要求非常高。媒体行业天然追逐新闻热点，假如知道一个事件是潜在的热点，媒体希望第一时间对热点进行追踪，报道和解读。媒体机构希望通过数芯平台，能够最快的发现互联网上发生的事情，发现热点之后以最快的方式，将结构化后的数据给到用户。

- 批流处理结合

基于媒体行业数据的特征，数芯平台采取批、流结合的方式解决目前客户和业务场景对数据的要求。批指的是大量的计算，基于平台特征完成复杂的模型，算法训练，长周期计算，文本实体识别，文本挖掘，借助批量的能力解决数据更深度，更大规模的加工。流指的是流式的计算，完成数据的清洗，结构化，轻计算和实时统计。出来一篇新闻，在整个新闻流传的过程中实时的将信息处理好。目前数芯整体的流计算大概需要300毫秒，即一篇新闻从源头过来，存储到平台，服务化出去中间需要经过300毫秒的时间。

- 批流结合的数据架构

数据从不同数据源过来，如API、OTS、ROS、log、file等，一方面数据要在实时计算当中做URL去重、正文结构化、标记来源、垃圾识别、实体识别和轻度的实时统计。另外基于MaxCompute，利用DataHub将数据存储到批处理平台中，由于流计算本身不做持久化存储，所以所有数据都将存储在MaxCompute上。数据存储之后做主题建设、关系挖掘、知识图谱计算、算法训练。通过批流结合的处理方式能够满足客户对数据本身能力的需要。之后，为用户提供搜索能力、大屏能力和BI能力等。

媒体大数据平台能力

- 内容结构化

人们在网页当中看到一篇新闻，而在数据库当中新闻是按字段存储的。比如，分为新闻标题，发布网站，时间，新闻来源，情感等等。平台需要将新闻信息进行结构化，成为后续计算过程需要依赖的数据结构化字段。

- 主题构建

媒体行业会分主题建设数据。平台会获取到不同的数据源，不同类型的数据。这些数据不可能完全的结合在一起，数芯平台会将所有数据分门别类的分成不同主题，按不同主题建设、存放和加工。媒体是一个非常复杂的行业，对各个行业的数据都有所诉求，媒体需要挖掘很多不同行业的数据支撑新闻生产和报道。目前，数芯集中在媒体内在的数据、体育、金融、气象等几个主题。数芯一方面将不同的数据源汇总到平台当中，另一方面数据进来之后，挖掘潜在的新闻点，生成选题策划，帮助用户做选题等工作。

- 实体识别

实体识别对于媒体大数据来说是最基础的能力。数芯目前积累的实体围绕人、机构、地点等三类数据。新闻行业，媒体行业会围绕某一个实体，关注与实体相关的数据能力。比如，很多公司会关注和自己相关的咨询，关心这些新闻是正面还是负面，哪些机构会发正面的信息，哪些机构会发负面的信息等。只有当将大批量的数据汇集起来之后，才能做相关内容的分析。实体识别场景在媒体大数据领域是非常基础的能力，首先，建立实体库。同时，当有一篇新闻产生，数芯需要实时的识别新闻与哪些人、机构和地点相关。另外，数芯需要汇集实体之间关系，做实体关系图谱。比如，很多品牌会瞄准自己的竞争对手，调整品牌战略，实体关系图谱对很多公司品牌运营的推广有很大的帮助。

- 情感分析

情感分析也是媒体大数据平台中比较通用的能力，当一篇新闻出来之后，用户需要知道是情感上褒义还是贬义。信息数量少的信息可以人工判断，但是如果每天有上千篇，上完篇的内容就无法通过人工来判断。媒体行业的情感分析与学术上的情感分析有差异。目前，自媒体出来之后短文本的内容越来越多。短文本的情感分析和长文本的情感分析不同，以往都是采用同一个算法实现情感分析，但发现效果并不好。现在，数芯将情感分析场景分开，微博短文本的情感分析用Word2vec+LSTM，新闻类的长文本的情感分析用Word2vec+CNN+RNN。分开之后发现每一类情感分析的效果都有所提升。

- 内容去重

内容去重是媒体大数据平台非常重要的一部分。去重能力是对于常见的新闻摘录，编辑，删减有准确的判重能力。一篇新闻并不是只由一个人写，它会被很多机构和渠道所转发。如何知道一篇新闻在哪些渠道被转发，其实就是通过去重的方式实现的。通过从大量的渠道汇集数据后，平台需要比对一篇新闻与之前哪篇新闻相似，通过相似度的比对得到结果。最早的时候，去重是基于关键词进行比对，数芯采取关键词和语义，两种方式比对，去重效果明显提升。内容去重可以用于新闻的热度计算，新闻数据顾虑清洗，文章版权追踪等业务场景。

- 内容标签化

搜索引擎可以用来搜索新闻，通过关键词和文章的匹配度来决定这篇内容是否推荐给用户。但纯搜索的方式已经满足不了用户家的需求。今日头条之所以成功，是因为它基于新闻和用户的习惯推荐内容。内容标签化就是通过机器的方式理解新闻，理解新闻与哪些信息相关，基于文本挖掘的手段，实现对于全网采集的内容数据进行分类打标。

Project 依赖

很多时候，使用大数据平台要不要分项目都是一个很难的抉择。不分项目的好处是开发人员都在同一个平台工作，互相之间不需要太多的授权，整体工作效率会比较高。分项目的好处是利用不同平台做不同的业务会更清晰，更具条理性。数芯在开始使用MaxCompute时，便采取分项目方式，其考虑的原因有以下三点。首先，分项目可以区分业务优先级，避免低优先级任务影响高优先级的数据产出。另外，可以区分资源消耗型，避免出现高资源消耗任务整体影响数据产出。还有区分内外服务，避免内部服务互相交叉影响。总体上，分项目可以为数据产出的稳定性提供很好的保障。

上云经验

- 由于媒体行业大部分数据都是非结构化的数据，会造成对单字段的容量要求比较大的问题。而且不同的平台和传输工具对于数据的字段大小的限制不同。在从不同平台做数据传输时，尤其需要关注这个问题。
- 能用UDF解决的问题，不要使用MR。使用UDF可以提高开发和运维的效率。即尽量用简单的表达式处理逻辑，这对整体数据产出稳定性有好处。
- 对查询效率要求不高的数据报表可以直连MaxCompute，减少中间环节。如此可以大大减少数据转换和数据维护成本。
- Datahub一方面可以接数据源，另外还可以较好的串联批，流之间的计算流程，保持数据一致性和形成依赖。
- 合理设计批，流的数据处理分工，减少重复计算。
- 媒体大数据经常需要运用不同的算法，PAI可以帮助解决很多算法问题，减少开发的工作量，提高数据处理的效率。

6.MORE VFX：阿里云桌面助力《流浪地球》视效艺术家“云复工”

电影里逼真、璀璨的特效背后，有着一群满腹才华的视效艺术家。你可能想象他们如梵高般疯狂、如李白般不羁，其实他们跟你一样，朝九晚五地上班。受疫情影响，许多公司选择在家办公，而制造、餐饮、零售、院线等线下行业则处于停摆。

其中，影视后期制作也是一个备受“复工难题”困扰的行业。许多读者印象中，影视后期制作的艺术家，都是聚精会神地坐在设备前，用各种酷炫的软件与技术做出或磅礴恢弘，或栩栩如生、纤毫毕现的特效。视效企业看似可以轻松实现在家办公，实际不然。国内顶级视效企业、《流浪地球》制作方MORE VFX与阿里云共同提出了一个解决方案，或为行业打开思路。



阻碍艺术家在家办公的几座大山

就拿《流浪地球》后期制作方MORE VFX为例，呈现视效行业的境况。

艺术家如果要在家里办公，继续完成自己的作品，可能需要破以下三道关：

- 首先是设备。建模和特效渲染等环节对CPU、显卡等电脑配置要求极高，特效公司一般会采购运算和实时图像处理能力更强的工作站——即定制的专业级电脑——给艺术家用。普通的个人电脑根本跑不起。这也是为什么很多艺术家们只能跟公司的猫一起加班，而无法享受在家办公的乐趣。即便将工作站从公司搬到了家里。艺术家可能会发现，都看完5部电影了，都还没下载完一个素材。
- 网络成了第二关。相比于其他行业，影视工业化生产过程中数据关联性很高且数据总量大，单个项目数据文件总量超过100TB。如使用家用网络下载到本地生产，根本是不切实际。
- 第三关则是数据安全。特效公司的素材是其核心资产之一，都存在公司内网中。艺术家在公司办公，本质上是在相对封闭的企业内网里工作，连电脑的USB接口都是封死的。要进入一家特效公司，你要经历的安保措施超乎想象，制作间更是“门派禁地”。艺术家如果要在家里办公，就要从公共网络连入企业内网，将尚未完成的作品数据重新加载，并且不断在内网检索查找已有的素材用于设计。如果素材泄漏，后果不堪设想；一旦素材涉及即将上映的电影，可能导致电影投资血本无归。

新冠肺炎疫情肆虐，严格隔离是保证员工健康的重要手段。然而，三座大山在前，特效公司的艺术家SOHO（Small Office, Home Office，在家办公）不易。视效制作公司需要人员协作密集生产，若项目无法正常调配，工资、房租等运营成本却在空耗，伤害是巨大的。

MORE VFX只花两天实现全员SOHO



如斯环境下，MORE VFX却意外迅速地实现了在家办公。据CEO徐建透露，他们从2月3日开始复工，目前云复工能保证日常生产70%-80%的进度，个别部门能达到90%，数据安全也得到了有力保障。

春节期间MORE VFX立刻针对员工健康成立疫情应对小组，制定防御措施。同时，“IT部门开始重新验证与阿里云合作的远程办公方案，当IT通过测试验收后，就由我们的CG总监先上手测试，并且逐步让每个部门都安排人员测试，以确定远程办公是否满足正常的生产流程。”徐建表示。

这里提到的远程办公生产方案，就是阿里云桌面/云图站。云桌面是一种基于桌面虚拟化的云服务。简单理解就是在云端服务器构建与本地电脑相同的操作环境，用户借助本地电脑屏幕对云端的电脑进行操作。

MORE VFX通过和阿里云数据中心拉通了专线，艺术家通过公网在家访问阿里云的云桌面实例，经远程连接软件（RDP、VNC、DCV等）跳板到办公室工作站。如此一来，艺术家真正在用的还是远方的电脑，本身就处于企业的内网环境。个人电脑再渣，只要网络正常即可，也不需要下载庞大的文件。

而在数据安全方面，作为跳板机的云桌面实例提供普通远程桌面工具（VNC、RDP等）所不具备的数据安全管控能力，降低MORE VFX的素材流失风险。比如可溯源的水印功能、可调整设备准入策略、禁用文件上传下载等。这些功能都让MORE VFX实现了“数据隔离”。

“第一天配置，第二天验证，第三天开始办公。”只用了两天，MORE VFX就实现了在家办公，恢复正常生产的同时保证了员工安全。有了一套相对完善的测试结果后，MORE VFX CEO徐建第一时间在社群里和行业二十多家公司做了技术分享，希望这次行业能共渡难关。

本次疫情，MORE VFX之所以能迅速恢复生产，得益于云的弹性。更不得不提的是，MORE VFX此前对云的积极拥抱，以及领先的技术储备，为后续办公业务的敏捷打下基础。

渲染环节先上云，引领行业技术升级



看似光鲜的背后，影视特效制作有着其传统之处。普通观众也许无法想象，在网络如此发达的今天，一部电影的渲染竟然是靠硬盘人肉运输到渲染农场的。据业内人士透露，一部电影，起码需要一百张硬盘。

被视作未来的云计算服务，在视效行业内尚待普及。幸而已有数家头部公司主动上云。MORE VFX就是其中一家引领者。去年起，MORE VFX就和阿里云达成了云的各方面合作。算力需求最大的渲染环节迈出了上云的第一步。MORE VFX后期特效量极大，同时对于计算力的需求随项目而变。其自有本地农场数百台多核服务器在低谷期利用率不高；在高峰期任务排队又严重，必须借助外部渲染农场服务器才能完成任务。而农场服务器老旧，性能不高；只能依靠硬盘寄送方式，资产安全性差；又缺乏弹性。

这些问题困扰着MORE VFX，最终它选择拉通上云专线，利用阿里云高性能计算服务（E-HPC），在云端完成渲染任务。在渲染任务上云过程中，结合后期制作公司的业务流程需求，阿里云团队提供了拉专线的混合云渲染方案，同时通过为MORE VFX定制文件异步缓存能力，有效降低了超大规模集群渲染时对专线带宽的压力。

此外，通过阿里云高性能计算缓存层的横向、纵向扩展能力，能够平滑分解渲染节点对缓存层的吞吐和IO压力，从而提供极强的资源弹性。通过阿里云高性能计算集成的行业通用软件deadline、计费 and 报表功能，渲染集群可以根据任务排队情况自动伸缩，使得运维极其简单，大大节省了MORE VFX IT部门的工作量。

视效行业空间巨大，技术红利有待激发

2019年下半年的渲染上云，MORE VFX提高了渲染性能与弹性，理解了云服务的便利。近期疫情，刺激了在家办公的强需求，MORE VFX不仅成为了远程云制作的开拓者，更是带动了许多影视视效企业开始关注云服务。

徐建曾经在接受文娱垂媒《三声》采访时表示：“我们跟好莱坞成功的视效公司相比，还存在几十年的差距。”他坦言，国外设计软件、研究底层算法；国内公司购买商业并使用软件，做少量二次开发。

MORE VFX主导或参与制作的影片多次获得香港金像奖等，代表作有《流浪地球》、《一出好戏》、《悟空传》等。作为国内一线电影视效公司，MORE VFX一直在探索如何引领并帮助行业转型升级。徐建一直期盼，国内产业能基础设施和核心算法上创新与突破，主动接触新技术，希望云计算能成为加速行业技术升级的平台。

徐建认为，目前中国的视效行业，不管是从市场需求或是整体发展来说，还有十年左右快速增长期。在互联网的上半场，云计算降低了互联网创业成本，提升创新效率；相信在产业互联网时代，云计算也将推动更多行业升级。

国内视效行业的发展前景巨大，技术红利仍有待激发。随着MORE VFX等行业上云，云计算服务影视行业的深度与广度将快速扩大，帮助国内视效行业构建属于自身的技术壁垒。

相关产品

● 云桌面

云桌面 (Cloud Desktop)，是由阿里云所提供的便捷、安全的云上虚拟桌面服务。它支持快速便捷的桌面环境创建、部署、统一管控与运维。无需前期传统硬件投资，帮您快速构建安全、高性能、低成本的桌面办公体系。可广泛应用于具有高数据安全管控、高性能计算等要求的金融、设计、视频、教育等领域。

更多关于云桌面的介绍，参见[云桌面产品详情页](#)。

● 弹性高性能计算

弹性高性能计算（E-HPC）基于阿里云基础设施，为用户提供一站式公共云HPC/AI平台服务，面向科研，生产，教育和行业大计算，提供快捷，弹性，安全和与阿里云产品互通的云超算平台。

更多关于弹性高性能计算的介绍，参见[弹性高性能计算详情页](#)。

相关最佳实践

[云桌面远程办公最佳实践](#)。

[E-HPC混合云渲染最佳实践](#)。

7.品推宝：营销内容与流程管理平台上云

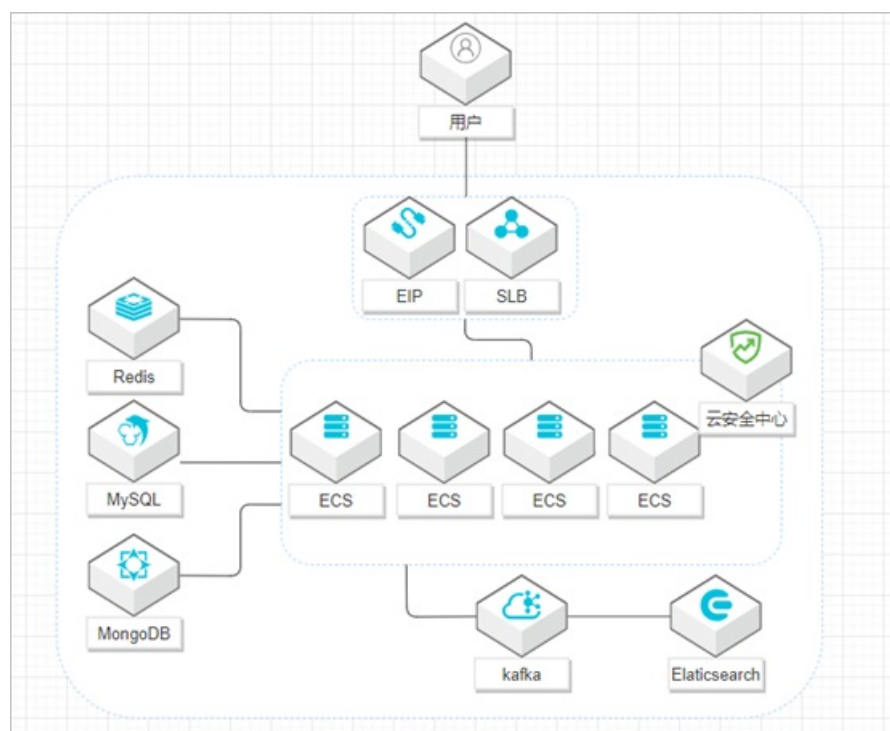
公司介绍

北京品推宝移动科技有限公司为品牌广告主或营销公司提供一个高度可视化、自动化、智能化的内容协同管理平台，用于大规模管理营销内容资产和委派内容创作任务。通过机器学习和推荐算法，实时分析海量数据来寻找、筛选、推荐与品牌营销目标相符合的个人或团队，从而帮助品牌广告主或营销公司找到最适合的合作伙伴和营销内容，将营销效果最大化。

业务痛点

- 我们公司打算新部署一个项目，对上线时间和研发效率要求高。对后期的业务规模不可控，要求设计弹性升级方案。
- 数据安全方面，希望有自动备份的机制，不需要我们的研发同学投入精力维护。
- 数据处理方面，针对多媒体数据具有强大的转码、压缩、剪辑等处理能力，最终输出为全平台支持的内容格式。
- 数据存储方面，希望具有安全、低成本、高可用、可靠性，并且容量和处理能力弹性扩展。

解决方案



方案细节：

- 使用阿里云的SLB配合EIP实现高可用架构，做流量分发，应用负载。
- 部署云安全中心实现服务器的基础防护，防御基本的木马病毒漏洞。
- 采用Redis做缓存数据库，解决活动期间的高并发访问瓶颈问题。
- 采用RDS存储结构化数据、用户信息等，RDS的高可用版以及自动备份策略可以满足我们公司数据的安全要求。
- 采用MongoDB存储非结构化数据，存储大量的BSON格式数据，实现数据分析。
- 使用Kafka配合Elasticsearch，做异步解耦，实现数据的检索以及日志的处理。

上云价值

利用这套架构，我们不仅实现了业务高可用，同时最大程度保障了数据的安全。配套产品节省了我们的开发周期，帮助我们快速搭建起了自己的应用体系。

证言

“感谢阿里云在我们上云期间为我们做的一切！响应很及时，服务很周到。”——北京品推宝移动科技有限公司 技术人员

相关产品

● 负载均衡SLB

负载均衡（Server Load Balancer）是将访问流量根据转发策略分发到后端多台云服务器（ECS实例）的流量分发控制服务。负载均衡扩展了应用的服务能力，增强了应用的可用性。

更多关于负载均衡SLB的介绍，参见[负载均衡产品详情页](#)。

● 云安全中心

云安全中心是一个实时识别、分析、预警安全威胁的统一安全管理系统，通过防勒索、防病毒、防篡改、合规检查等安全能力，帮助用户实现威胁检测、响应、溯源的自动化安全运营闭环，保护云上资产和本地主机并满足监管合规要求。

更多关于云安全中心的介绍，参见[云安全中心产品详情页](#)。

● 云数据库 Redis 版

阿里云数据库Redis版是兼容开源Redis协议标准、提供内存加硬盘混合存储的数据库服务，基于高可靠双机热备架构及可平滑扩展的集群架构，可充分满足高吞吐、低延迟及弹性变配的业务需求。

更多关于云数据库 Redis 版的介绍，参见[云数据库 Redis 版产品详情页](#)。

● 云数据库MongoDB版

云数据库MongoDB版支持ReplicaSet和Sharding两种部署架构，具备安全审计，时间点备份等多项企业能力。在互联网、物联网、游戏、金融等领域被广泛采用。

更多关于云数据库MongoDB版的介绍，参见[云数据库MongoDB版产品详情页](#)。

● 消息队列 Kafka 版

消息队列 Kafka 版是阿里云基于 Apache Kafka 构建的高吞吐量、高可扩展性的分布式消息队列服务，广泛用于日志收集、监控数据聚合、流式数据处理、在线和离线分析等，是大数据生态中不可或缺的产品之一，阿里云提供全托管服务，用户无需部署运维，更专业、更可靠、更安全。

更多关于消息队列 Kafka 版的介绍，参见[消息队列 Kafka 版产品详情页](#)。

● 阿里云·Elasticsearch

提供100%兼容开源Elasticsearch的功能，以及Security、Machine Learning、Graph、APM等商业功能，致力于数据分析、数据搜索等场景服务。

更多关于阿里云 Elasticsearch的介绍，参见[阿里云·Elasticsearch 产品详情页](#)。

● 弹性公网IP（EIP）

独立的公网IP资源，可以绑定到阿里云专有网络VPC类型的ECS、NAT网关、私网负载均衡SLB上，并可以动态解绑，实现公网IP和ECS、NAT网关、SLB的解耦，满足灵活管理的要求。

更多关于弹性公网IP（EIP）的介绍，参见[弹性公网IP（EIP）产品详情页](#)。

● 云服务器ECS

云服务器（Elastic Compute Service，简称ECS）是阿里云提供的性能卓越、稳定可靠、弹性扩展的IaaS（Infrastructure as a Service）级别云计算服务。云服务器ECS免去了您采购IT硬件的前期准备，让您像使用水、电、天然气等公共资源一样便捷、高效地使用服务器，实现计算资源的即开即用和弹性伸缩。阿里云ECS持续提供创新型服务器，解决多种业务需求，助力您的业务发展。

更多关于云服务器ECS的介绍，参见[云服务器ECS产品详情页](#)。

- 云数据库RDS

阿里云关系型数据库（Relational Database Service，简称RDS）是一种稳定可靠、可弹性伸缩的在线数据库服务。基于阿里云分布式文件系统和SSD盘高性能存储，RDS支持MySQL、SQL Server、PostgreSQL、PPAS（Postgre Plus Advanced Server，高度兼容Oracle数据库）和MariaDB TX引擎，并且提供了容灾、备份、恢复、监控、迁移等方面的全套解决方案，彻底解决数据库运维的烦恼。

更多关于云数据库RDS的介绍，参见[云数据库RDS产品文档](#)。

8.南瓜电影：实现移动端与服务器双向互通

公司介绍

南瓜电影App是国内领先的专注于影视精品化运营的垂直类视频产品，在移动互联网、OTT等客户端为用户提供差异化内容运营服务，是国内唯一为用户提供专注于精品电影电视剧的全会员制视频App。南瓜电影曾出品大量制作精良、题材健康的优质自制剧及电影，并与全球超过150家独立制片公司达成战略合作，获得优质影视内容版权。经过多年的发展，南瓜电影在国内为超过2千万的会员提供优秀精品影视服务，还在海外特别是东南亚地区积累了大量用户群体。

业务痛点

在南瓜电影的业务场景中，用户客户端与服务端之间存在频繁的双向交互，比如续费管理、内容推送、会员互动、评论提交等，其中有很多业务消息都需要同时发往多个客户端。

- 传统的Websocket协议或者基于TCP/UDP自建通讯协议都存在业务逻辑实现困难。
- 处理断线重连，错误重发等复杂的技术问题导致开发负担重，对用户体验好感度造成巨大影响。

解决方案

为了适应业务的快速发展，提升用户体验，南瓜电影决定采用MQTT方案来解决服务端与客户端之间的双向消息通讯。

- 阿里云提供的微消息队列MQTT+消息队列RocketMQ产品组合非常完美的实现了这个方案，让南瓜电影的技术团队通过非常简单的方式，快速接入这一套成熟、健壮的MQTT消息体系，减少了处理种类底层技术问题所耗费的精力。
- 对于海外的用户群体，阿里云消息队列RocketMQ提供的全球路由功能能够实现跨地域之间的毫秒级消息同步，配合客户端的MQTT接入，能使海外用户的操作更为流畅，在用户体验方面带来了极大提升。

上云价值

- 使用微消息队列MQTT+RocketMQ的组合，用最简单的方式实现了海量移动端App和服务端之间的双向互通，而且这种双向互通是直接具备异步分发以及负载均衡属性的，不需要另外再使用其他的负载均衡以及服务端消息分发应用组件。
- 另外，通过RocketMQ的全球路由机制，能够让海外的客户端与本地的服务端之间实现毫秒级别的消息传输，既确保了消息的可靠性，也确保了消息的低延迟。

证言

“引入阿里云的MQTT方案后，南瓜电影在用户体验方面得到了很大提升，IT团队也更有信心地应对用户量的快速增长。”——南瓜电影CTO 庄徐麟

相关产品

● 微消息队列 MQTT 版

微消息队列 MQTT 版是专为移动互联网(MI)、物联网(IoT)领域设计的消息产品，覆盖互动直播、金融支付、智能餐饮、即时聊天、移动 Apps、智能设备、车联网等多种应用场景；通过对 MQTT、WebSocket 等协议的全面支持，连接端和云之间的双向通信，实现 C2C、C2B、B2C 等业务场景之间的消息通信，可支撑千万级设备与消息并发，真正做到万物互联。

更多微消息队列 MQTT 版的介绍，参见[微消息队列 MQTT 版](#)。

- 消息队列 RocketMQ 版

消息队列 RocketMQ 版是阿里云基于 Apache RocketMQ 构建的低延迟、高并发、高可用、高可靠的分布式消息中间件。该产品最初由阿里巴巴自研并捐赠给 Apache 基金会，服务于阿里集团 13 年，覆盖全集团所有业务。作为双十一交易核心链路的官方指定产品，支撑千万级并发、万亿级数据洪峰，历年刷新全球最大的交易消息流转记录。

更多关于消息队列 RocketMQ 版的介绍，参见[消息队列 RocketMQ 版产品详情页](#)。