

12（二） | 数据被加工后，你还要学会使用和管理数据

2020-05-04 郭忆

数据中台实战课

[进入课程 >](#)



讲述：郭忆

时长 11:43 大小 10.73M



你好，我是郭忆。

上一讲，我讲了数据研发的四个阶段，你可以发现，标准化的研发流程对交付高效、高质量的数据来说，非常关键。那么数据被加工好以后，怎么使用数据和管理数据就是重点了。

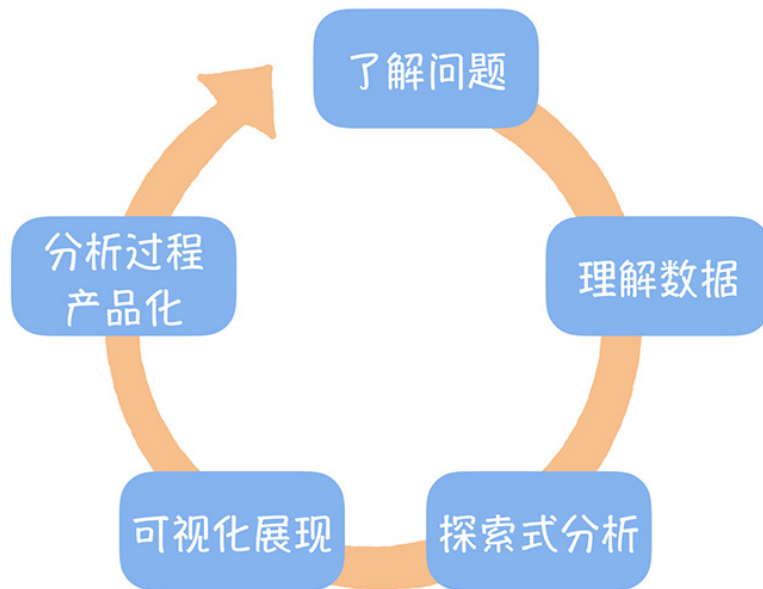
所以今天，我会从数据使用者的角度出发，聊一聊怎么构建高效的数据分析流程。同时，也会以资产管理者的视角，带你了解怎么实现数据资产的精细化管理。



我希望你通过学习今天的内容，判断一下，日常工作中自己在数据使用和管理方面是不是还存在流程环节上的缺失，并不断完善，让数据使用、管理得更好。

落地高效的数据分析流程

根据我的经验，我把数据分析过程划分五个步骤。接下来，我通过分析师甄可爱的例子，为你呈现了一个典型的数据分析流程。



第一步：发现业务问题。

数据分析的典型场景呢，起点都是业务出现了某个问题，我们需要基于数据找出业务问题背后的原因。

分析师甄可爱所在的公司，电商平台 Q2 季度某个品类的商品销售额下降了 30%，老板要求给出问题的原因，并进行整改。这个任务落到了她的身上。要解释这个问题，她必须从现有的数据入手，看看到底是哪里出现问题。

第二步：理解数据。

她首先要了解这样几点：

要分析的业务过程；

这些业务过程中涉及到了哪些关键指标；

这些指标的业务口径是什么；

有哪些可以分析的维度。

这些事儿，比较琐碎，甄可爱为了提高效率，利用指标系统，将要分析的业务过程快速锁定到交易域下的业务过程，然后找到交易域下有哪些指标。通过指标系统，她了解了“渠道销售额”这个指标的口径定义、计算逻辑和数据来源。

接下来，她要去查看指标对应的数据，借助指标系统，甄可爱可以直接跳转到指标关联到数据报表上，接下来她需要申请报表的权限，查看数据。报表负责人审批通过后，甄可爱就可以看到数据了。

数据查询

数仓表索引

主题域(13)

全部主题域 (系统)

商品域

采购域

品控域

仓储域

交易域

营销域

售后域

物流域

流量域

财务域

用户互动域

外域

质量域

全部主题域

请输入表名称

数仓表

表分层: 全部 dim dwd dws ods dm mid

所属主题域: 全部 商品域 采购域 品控域 仓储域 交易域 营销域 售后域 物流域 流量域 财务域 展开

更多筛选 重置

#	库表名称	表负责人	表描述	主键	表分层	主题归属	操作
1	dim.dim_access_disabtnfo		dis_abt 平台abtest配置信息	-	dim	流量域	查看核心属性
2	dim.dim_anonymous_user		匿名用户	-	dim	用户互动域	查看核心属性
3	dim.dim_author_sku		SKU行权限维表	-	dim	财务域	查看核心属性
4	dim.dim_author_spu		SPU行权限维表	-	dim	财务域	查看核心属性
5	dim.dim_bonus_activity		津贴活动维表	-	dim	营销域	查看核心属性
6	dim.dim_bonus_feedback_activity		回馈金活动维表	-	dim	营销域	查看核心属性
7	dim.dim_bonus_feedback_k_obtain		回馈金获取维表	-	dim	营销域	查看核心属性
8	dim.dim_bonus_obtain		津贴获取维表	-	dim	营销域	查看核心属性

数据地图导览示意图

这个时候，她发现，淘宝渠道销售额数据出现下降，拖累了整体品类销售额的数据。可是当她想进一步探查渠道下降的原因时，却发现并没有渠道级别的商品库存和销售指标。现在，靠现有的指标和数据已经没办法进一步解读业务问题的原因了，甄可爱需要进行探索式分析。

第三步：探索式分析。

那她首先要找到当下有哪些数据可以用，借助数据地图， she可以快速了解当前主题域下有哪些表，这些表分别代表什么含义。

这个时候，会存在两种情况：

如果现有的数据可以满足分析的需求，她可以直接在数据地图表详情页上发起数据权限的申请流程；

如果现有的数据没办法满足需求，甄可爱就要对数据开发提出数据研发的需求，会稍显麻烦。

幸运的是，甄可爱发现，商品粒度的库存和销售表中有渠道的字段，按照渠道进行聚合、过滤，就可以满足分析的需求了。所以，她在数据地图的相关表详情页里申请了这些表的权限。

接下来，权限申请流程会流转 to 表对应的负责人上：

对于核心表（比如交易数据），除了表负责人审批，还需要中台负责人审批；

核心表中的一些核心 KPI 数据（比如平台全年销售额），还需要 CTO 甚至 CEO 级别的审批。

等了一段时间，权限审批终于通过，甄可爱收到了来自权限中心的通知，于是她马不停蹄地在自助分析上，基于 SQL 对相关表进行了探查分析。甄可爱对比分析后发现，淘宝渠道销售数据下降的主要原因是：该品类下的部分畅销商品经常库存为 0，出现缺货情况，导致整体品类销售额下降。

第四步：可视化展现。

现在，找到了问题原因，为了给老板讲清楚分析过程，甄可爱还要通过报表的方式，把分析过程呈现出来。所以，她又在 BI 工具网易有数上进行了报表的制作，把报表授权给相关的管理层。

看到了原因后，管理层制订了供应链优化措施，加大了淘宝渠道的库存供货，整体品类销售额数据出现回升，终于解决了问题。

第五步：分析过程产品化。

解决了现有问题，并不是数据分析的终点。我们还要建立长久的问题发现和解决机制。

为了持续地监控该问题，并对其进行智能预警，甄可爱需要将分析过程固化到数据产品中。她策划并研发了供应链决策协同系统，能够自动检测商品的库存和销售，智能生成补货建议，然后推送给采购系统。

到此，整个数据分析的全过程就完成了。最后，我想再强调一个点，在这五个步骤中，你往往最容易忽略是最后一个步骤。当然，这也并不只是分析师的疏忽，本身数据产品的建设还需要有一定的研发资源的投入。

为了解决大规模数据产品研发资源投入的问题，在网易，我们基于网易有数（BI 工具）实现了数据门户的功能，它实现了一个低代码构建数据产品的开发环境，允许分析师通过拖拉拽的方式构建企业数据门户，从而为高效的大规模数据产品构建提供了基础。基于数据门户，企业可以构建商品运营系统、供应链协同决策系统、流量看板系统、会员运营管理系统等不同的数据产品，满足不同场景下数据分析的需要。

数据如何被使用讲完，接下来，我还想来谈谈数据的精细化管理流程，因为这个流程或者环节的缺失，会导致很多成本、安全、以及稳定性的问题。

构建精细化的资产管理流程

在数据中台中，数据资产的精细化管理主要包括成本治理和资产管理两个部分。在网易，我们分别研发了两个工具产品来完成上述管理流程的落地，分别是成本治理中心（简称 EasyCost）和数据管理中心（简称 EasyManager）。

下面我们通过资产管理李无邪的视角，来看看上述两个工具产品日常是如何运转的。

李无邪首先要登录到 EasyCost 中，然后制订数据自动下线的规则，比如，他认定 30 天内没有访问的数据需要下线。然后系统会根据规则，每天自动将符合规则的表和目录推送给表的负责人，等待表的负责人审核确认。

表的负责人张美丽接到了 EasyCost 推送的邮件，此时一般有两种情况：

第一种，是该数据虽然没有被使用，但是属于核心资产，以后用的上，需要保留，此时可以申请加入白名单中，由资产管理李无邪审批后，不再被推送。

第二种情况，是该数据确实没有被使用了，那张美丽就点击一键下线，然后系统会进行数据的灰度下线，首先会先停止调度任务，数据不再产出，7 天后，数据会被自动清理。在下线前，可以选择是否保存备份。

为主题域的负责人和数据团队的管理者，同样也会收到 EasyCost 推送的面向主题域和数据中台整体的表的使用情况，从管理者的角度，也可以对下形成治理的压力，把成本治理纳入到数据开发的绩效考核中。

接下来，我们讲讲资产管理部分。资产管理的核心，是数据资产等级的制订，李无邪需要为数据中台的数据制订资产等级规则。

李无邪要依据两方面的因素，制订资产等级的标记规则：

一方面是数据本身涉及企业的核心机密，比如 KPI，产品日活，毛利等；

另外一方面因素是根据数据应用的优先级，然后基于全链路的数据血缘制订数据的等级。

数据等级可以与数据权限的审批流程、模型和任务发布上线的审批流程打通，根据不同的资产等级，需要不同级别的角色来完成审批。另外，数据资产等级还与数据备份策略相关，对于核心数据，我们要求必须实施备份。

此外，数据中台的小文件也需要关注，因为如果小文件过多，会导致 HDFS 元数据过大，对 HDFS 的元数据服务 NameNode 产生性能问题。所以 EasyManager 同样需要对小文件的数量和分布进行监控，然后推送给各个主题域和表的负责人，同时系统提供了小文件合并的工具，可以帮助数据开发快速的完成小文件的治理。

课堂小结

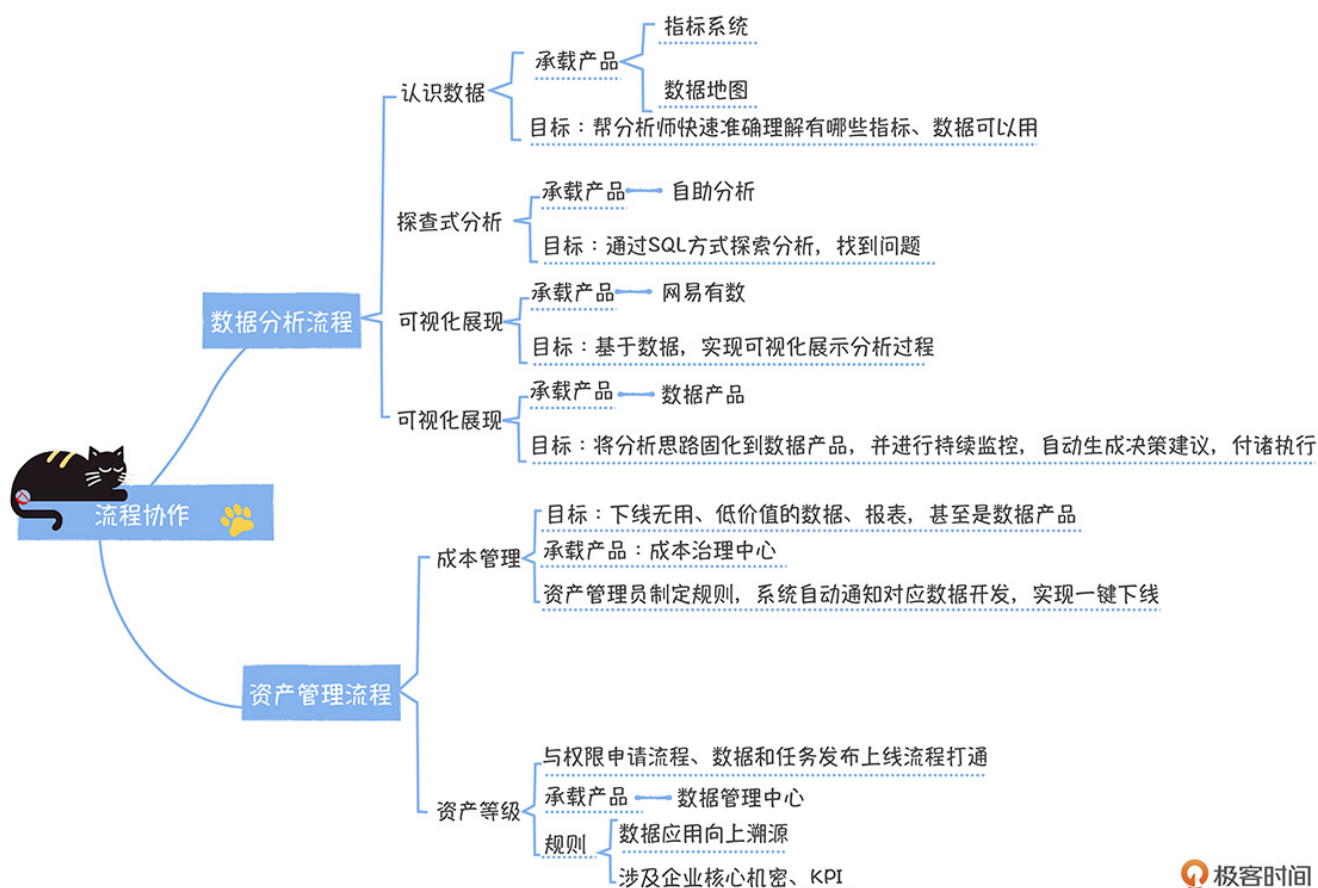
今天这节课，我带你重点了解了如何构建高效的数据分析流程，和如何实现精细化的资产管理流程。

通过这两讲内容的学习，我相信你就不会觉得，面对这么多的工具产品，不知道该怎么用！涉及这么多人，又不知道什么人该干什么事儿了。同样，你也可以把前面提到的工具和角色串联起来，形成一个可落地运行的机制，应用到你日常的数据建设工作中。

在最后，我想再强调几个重点：

数据分析的完整流程应该从了解业务数据，到探查式分析，再到通过数据报表进行可视化呈现，最后通过数据产品固化场景，实现持续监控、自动生成决策建议，一键执行的目标。

资产管理流程中，资产管理员的主要职责在于制订规则，包括数据或者报表下线的规则，数据资产等级的规则，目的是凸显数据的资产属性，聚焦核心数据。



思考时间

数据研发、数据分析以及资产管理是数据中台中三个基本流程，除了这些，你还知道有哪些别的流程需要涉及到多个角色的协作？如果需要通过一个工具产品，流程协作中心来完成上述协作流程，你觉得该如何设计这个产品呢？

欢迎在留言区与我互动。最后，感谢你的阅读，如果这节课让你有所收获，也欢迎你将它分享给更多的朋友。

五一计划

晒学习姿势 「免费」领课程



【点击】图片, 立即参加 >>>

© 版权归极客邦科技所有, 未经许可不得传播售卖。页面已增加防盗追踪, 如有侵权极客邦将依法追究其法律责任。

上一篇 12 (一) | 数据研发就只是写代码吗?

精选留言 (1)

写留言



吴科

2020-05-04

五一又更新了!

除了数据研发, 数据分析, 资产管理。应该还有营销系统, 数据中台输出分析数据给营销使用, 推送营销短信, 优惠券, 客服打电话等。

数据只有与业务关联后才有意义, 业务提需求, 数据研发上线, 营销推广。涉及产品经理, 分析人员, 数据开发, 前端开发, 营销人员多个角色。这里沟通最重要, 数据指标...

展开



