

13 | 数据中台在网易电商业务的最佳实践

2020-05-06 郭忆

数据中台实战课

[进入课程 >](#)



讲述：郭忆

时长 19:15 大小 17.64M



你好，我是郭忆。

从 OneData 到 OneService、从方法论到支撑技术、从宏观架构到微观实现，我已经带你系统地了解了建设数据中台的所有核心知识点。你可以看到，虽然人人都在讨论数据中台，但是真的实战起来，还是不简单吧？

而且建数据中台是一项系统性的工程，涉及人员组织架构的变动，需要研发大量的系统支撑工具，更要和业务部门达成密切的合作，形成双赢，反之会有失败的风险。还是分享- ☆ 见过的的事儿。

甄英俊是某零售企业 IT 部门的老大，最近他也想在企业中建数据中台。设想一番后，他亲自操刀，组建了新的数据中台部门，还亲自规划了十个业务场景（包括会员看板、商品运营、供应链管理、售后管理、毛利分析、类目管理、门店管理、仓储管理、渠道分析、辅助选品）。

但数据中台团队没有和业务部门达成一致的 KPI，在具体工作推进过程中，中台团队与业务部门脱节，业务部门也没有资源支撑中台的推进（例如指标的梳理）。

最后，虽然基于原先规划的十个场景，数据中台确实做出了一些报表，但很少有人查看。于是，尴尬的一幕发生了：在年终总结汇报中，甄英俊自信地向 CEO 汇报了数据建设的成果（输出了多个报表，覆盖了多少业务场景）。可当 CEO 问业务老大是否感觉到数据的作用？业务老大摇了摇头，他们并没有认可数据中台的成果。

这是一个很典型的失败项目，而问题的根源就在于数据中台团队虽然独立于业务，但是并不能脱离业务。甄英俊最大的失误就是没有深入调研业务问题，也没有和业务达成一致的 KPI，更没有根据业务的反馈，不断完善数据应用。

所以，如果你要建中台，要做到这样几点：

问问自己为什么要建中台，与业务达成一致的目标；

把数据中台作为一个公司级别的顶级项目来推进，而不是一个数据部门自己的 KPI；

数据中台必须要有清晰的、可量化的价值来衡量（从主观上也要得到业务部门的认可）。

而今天这节课，我会从项目立项、到项目推进，最后到项目总结，带你看一下网易电商数据中台的构建过程。这样一来，你会明白如何在企业一步一步落地数据中台，这对想要建设数据中台的你说来，非常有借鉴意义。

立项数据中台项目

我认为，立项是建数据中台最关键的一步，因为它的核心就是挖掘业务的痛点，跟业务达成一致的建设目标。**如果能达成一个一致的、可量化的目标，数据中台的项目就成功了一半。**

所以在网易电商数据中台构建之前，我们（数据部门）对业务方（包括市场、商品、供应链、仓配、售后、会员等）进行了密集的调研，尤其是跟各个部门的老大了解了这样两个方面：

当前数据使用过程中存在哪些痛点；

当前业务部门最关注的业绩目标。

这里我多说几句，对一些传统企业来说，业务部门的数据思维能力比较薄弱，数据使用水平还比较初级，根本讲不出什么痛点。如果遇到这种情况，你要多关注一下业绩目标（比如，如何让数据帮助企业达成 KPI。）如果谈论这种话题，业务部门的老大一定很感兴趣。

经过调研，我总结了这样几个痛点。

第一，指标业务口径不一致。

在建数据中台前，网易电商已经有 20 多款数据产品，它们涉及 800 多个指标，存在大量业务口径不一致的情况，导致了数据不可信、无法用。你看，虽然数据产品不少，但是真正发挥价值的却寥寥无几。其实这还没算数据报表，如果把报表算进来，就更夸张了。

第二，需求响应速度慢。

在电商平台中，业务部门运营活动的规模和频率大大提高，原先可能一个月才一次促销活动，现在是一天一次，甚至还有小时级别的。淘宝上一双 NewBalance 的鞋子，前一个小时还是 599，下一个小时就成了 429，而这背后其实是大量的商品运营策略在发挥作用。

活动一开始，运营人员就需要分析大量的数据，从而不断优化运营策略。此时，数据部门会收到大量的数据研发需求。而我们统计后发现，数据部门一个需求的平均交付时间是一周（数据来源于项目管理工具 JIRA），根本没办法满足业务部门的需求，可以这么说，数据研发的速度已经无法支撑业务高频的运营活动。

第三，取数效率低。

我们问了很多的分析师、运营，他们集中认为取数效率太低，原因有两个。

一个是他们不知道有哪些数据，也不知道到哪里去找数据。当时整个电商团队存在三个小数仓（供应链、市场和仓配客）加起来有近 4W 张表，对他们来说，找到并准确理解数据，其实是非常困难的事情。

第二，基于 SQL 取数，对于非技术人员来说，门槛比较高。分析师经常遇到一个 SQL 异常就不知所措，更不要说不懂 SQL 的运营。

正是因为这两个原因，取数要靠数据开发帮助完成，效率很低。有多低呢？平均取数需求从提出到交付，需要一周（数据来源于项目管理工具 JIRA）。而这也抑制了数据的大规模使用，与此同时，临时取数的需求，占据了数据开发的大量时间（来自 JIRA 的数据统计，数据开发 50% 的时间都被临时性的取数需求占据）。

第四，数据经常违反常识。

糟糕的数据质量也是各个业务部门对数据最为不满的地方，经过 POPO 群统计（网易内部办公协作通讯工具），平均每周，我们就有 10 个数据质量问题被业务方投诉。

更为可怕的是，这些问题中，90% 都是由业务方先于数据提供方发现的问题，然后 50% 都是因为数据研发的 BUG 导致。在当时，我们经常出现数据经常无法按时产出，数据修复需要花费一天的时间！

第五，数据成本指数级增长。

2018 年，电商业务的大数据资源从一年 4000CU(1 CU = 1 Core + 4 memory) 增长到 12000CU，并且还在持续高速增长。当时正好是 2018 年底，公司对业务毛利这块儿管控的非常严格，精简成本作为了各个业务推进的优先级，大数据这块也不例外。

为了优化大数据的成本，我们排查后发现，至少有 20% 的数据在当时都属于废弃数据，这些数据每天仍在产生，却没有人访问，浪费着资源。

除了现有数据是否用的好以外，我们也对各个部门的业务目标进行了调研，目的就是让数据帮助解决更多的业务问题。

商品部门：主要目标是优化商品结构、降低滞销商品比例、提高商品库存周转，从而达到控制毛利水平的目标。所以他们最紧急的就是监控平台上滞销的商品。

供应链部门：主要目标是尽可能保证商品的供货充足，尽量避免缺货商品出现。所以及时发现缺货商品，制订更精准的采购计划是最紧急的事儿。

仓配客部门：最重要的业务目标是保障商品及时送达，优化物流成本。所以，基于各个仓库的数据和物流公司的报价，制订最合理的配送计划，是他们最重要的事儿。

有了这两方面的调研之后，我们下一步的目标制订就会顺利很多，最后，我们从效率、质量和成本三个方面，和业务部门制订了共同的 KPI。同时又因为当时整个电商业务的核心方向是控制毛利水平，提高商品周转，所以在业务目标上，我们选择了与之最相关的商品和供应链部门进行合作，共背业绩 KPI。

这里我提供给你一份数据中台 KPI 考核表格，希望你能参考一下，数据中台部门考核 KPI 应该包括哪些内容。

大模块	子模块	比例	内容	目标值	评分方法
中台建设	效率	30%	1. 研发效率提升，通过模型复用，数据服务化，提升数据的开发和交付效率。 2. 自助取数，提升取数效率，释放取数需求。	1. 模型复用比提升到3以上（目前1.7） 2. 平均需求交付时间缩短50%（当前一周） 3. 自助取数比例占50%（当前每周100次取数工单）	按照完成比例折算
	质量	10%	1. 建立全链路监控，通过稽核监控，提前发现问题。 2. 梳理指标，建立指标管理体系，消除数据产品口径不一致问题。	1. 数据质量问题下降50%（当前每周10个） 2. 100%消除数据产品指标口径不一致问题。	按比例折算
	成本	10%	清理无访问的表，消灭低价值资产	节省20%的成本。（优化的任务/表资源消耗占当日数仓总成本的比例）	按完成比例折算
业务支撑	商品	30%	建立商品动销、滞销监控体系，提前发现问题商品	滞销商品比例降低50%	按完成比例折算
	供应链	20%	智能预测商品缺货风险，自动发起商品补货	系统发起的商品补货占整体采购订单50%	按完成比例折算

你看，这个表里包含中台建设和业务支撑两部分，前者对应的是业务痛点，后者对应的是业务目标。更为关键的是，我们都是从业务出发制订的这两部分内容，我认为这是业务愿意和中台团队达成共建 KPI 的基础。

后来，在 CTO 的推动下，供应链、仓配以及市场部门把指标梳理、自助取数、数据模型迁移中台纳入了 KPI 考核。当然，对数据中台的支撑工作，这部分在业务部门的 KPI 中比例不会很高，一般最多 20%，但是却很重要，因为只有这样，业务部门才有压力去做这个事情。

以上就是立项阶段了，在这个阶段里，我们从挖掘业务痛点，到输出共建目标，接下来，我们就来看一看这个目标是怎么落地的。

推进数据中台项目落地

我会从四个方面带你了解数据中台的落地过程，这部分内容，会帮你把前面 12 讲串起来，并形成企业落地执行方案。

第一步，调整团队组织架构，明确各个团队的职责。

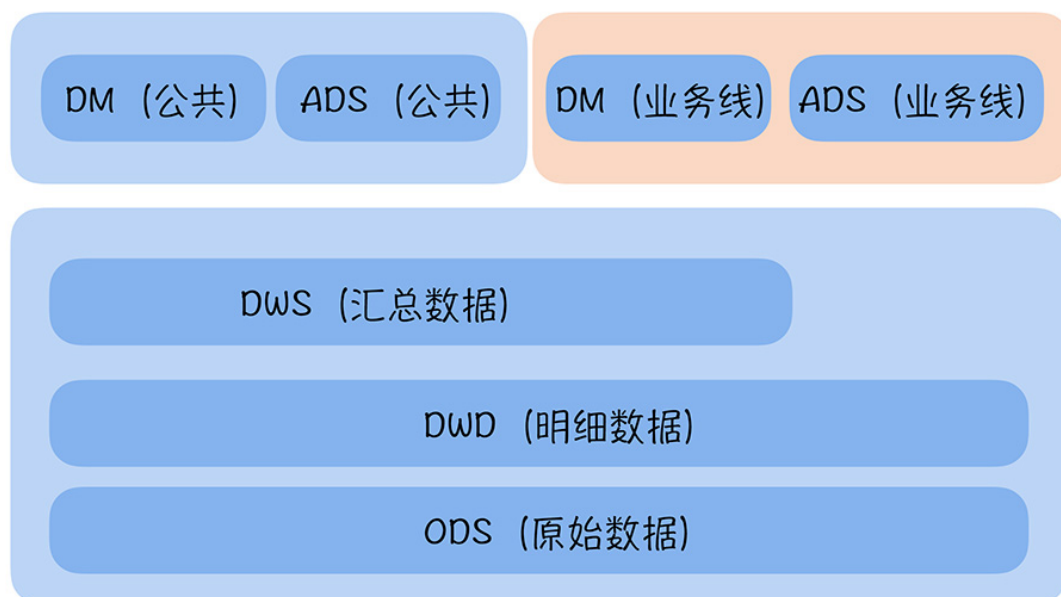
在数据中台构建之前，电商业务主要存在 3 个独立的数仓：市场、供应链和仓配客。这些业务部门中有数据开发、数据产品还有分析师。

而我们首先要做的，就是成立数据中台团队，这个团队是在原有市场数仓（除了服务市场，还服务于管理层）的基础上建起来的。

而对供应链和仓配客，我们并没有立即把他们的数据开发团队并入中台团队，而是调整了团队职责，把他们的数据开发的职责调整成，基于数据中台数据，加工私有的集市层和应用层。

这样处理的话，各方比较容易接受，不然业务部门会觉得中台团队在抢他们的人，对于员工个人，也可能因为团队定位、福利等原因不愿意转部门。

建立数据中台团队之后（主要由 3 名数据产品和 15 个数据开发构成），接下来就是明确团队的职责。数据中台主要负责 DW 层公共数据，以及跨部门共享的集市层和应用层的数据建设。



数据中台团队与业务部门分工示意图

第二步，数据整合。

中台团队成立后，首先面对的是混乱的指标业务口径，所以团队要先梳理指标，建立全局的指标管理规范（对应咱们的 05 讲）。这项工作由 1 名数据产品牵头，2 名数据产品辅助，对电商分布在各个业务线的 20 多个数据产品，800 多个指标进行了梳理，去除了冗余指标，对齐口径不一致的指标。

最后，我们把指标梳理成 400 个，其中，原子指标 127 个，这个过程花了 1 个月的时间，不过大部分时间都是在理解业务，和业务的分析师、数据产品对口径。

接下来，中台团队还要对模型进行重构、整合和迁移（对应咱们的 06 讲），而这部分工作可以分为设计阶段和实施阶段。

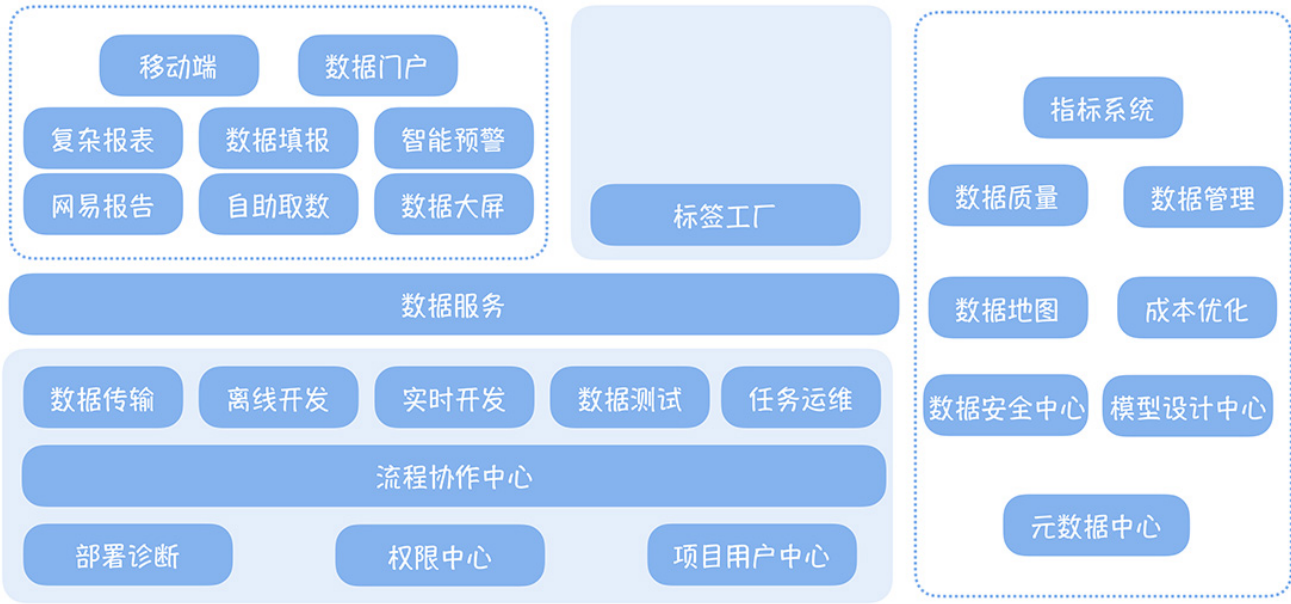
设计阶段，由一名数据架构师牵头，根据梳理好的指标，构建主题域，然后在原有模型的基础上进行重新设计，构建一致性维度。**这里需要强调的是**，中台团队必须要完全接管 ODS 层数据，这可以强迫业务部门必须要基于中台数据进行再加工。当然，中台团队会肩负着巨大的压力，但是只有熬过最痛苦的时期，这种中台的机制才能建立起来，**一旦原始数据没有管住，那中台就会功亏一篑。**

实施阶段需要投入大量的人力，但是中台团队还承担着公共数据需求的研发，所以为了保证模型重构和迁移的进度，我们从 15 个数据开发中拆出了 5 个人，专门进行模型的重构和迁

移。最终，我们完成了 200 多张表的基础数据的迁移重构，而这个过程花费了近 5 个月的时间。

第三步，研发工具产品。

在数据中台构建过程中，我们积累了很多规范和经验，但数据中台如果要形成落地、长久的运行机制，就必须把这些规范和经验沉淀到产品中，通过产品化的方式实现。



网易数据中台支撑技术工具产品架构图

所以在原有数据研发、数据产品团队的基础上，我们开始构思数据平台（工具产品）研发团队。因为考虑到网易集团存在公共技术研发部门（杭州研究院），可以实现工具产品在集团内不同业务线之间复用，所以选择了与公技合作的方式，由公技承担数据中台支撑技术产品的研发。

我们研发了 20 多个数据中台支撑技术产品，总结了四个产品设计的经验，对你设计这些产品有很大的借鉴价值。

正交化产品设计。每个产品聚焦一个应用场景，构建产品矩阵，简化了系统使用的复杂度。

类别	产品	英文名称	核心功能	使用对象
数据研发	数据传输中心	EasyTransfer	提供各种数据源的离线和实时数据传输	数据开发
	离线开发中心	EasyDev	基于Hive、Spark的离线数据开发平台	数据开发
	实时开发中心	EasyFlink	基于Flink的一站式实时数据开发平台	数据开发
	数据测试中心	EasyTest	提供数据比对、静态代码检查的数据测试工具	数据开发
	任务运维中心	EasyOps	提供运维大屏、全链路影响分析、任务管理（重跑、补数据）、智能报警、任务治理等功能	数据开发
	流程协作中心	EasyCoordinator	多角色流程协作，包括审批、通知等	数据开发、应用开发、资产管理 员、数据产品、运营、分析师
数据治理	数据地图	EasyDmap	企业元数据门户，查询有哪些数据	数据开发、分析师
	指标系统	EasyIndex	统一管理指标的业务口径定义，消除指标业务口径不一致，形成企业的指标字典。	数据产品（管理指标）、分析 师、运营、数据开发、应用开发
	数据质量中心	EasyDQC	稽核校验，全链路数据质量监控	数据开发
	成本优化中心	EasyCost	消除无用的、低价值的数据，建立数据的ROI评估体系	资产管理、数据开发
	数仓设计中心	EasyDesign	模型设计开发平台，基于维度建模理论，通过构建主题域、分层的方式组织数据，基于指标、维度、度量构建数据模型。	数据开发
	数据管理中心	EasyManager	数据备份与恢复、数据生命周期管理、文件管理、数据资产等级	资产管理、数据开发
	数据安全中心	EasySecurity	提供数据脱敏、数据加密功能	数据开发
数据服务	数据服务	EasyDS	提供统一的API管理平台，基于数据中台的数据，一键发布API接口，提供类型丰富的中间存储	数据开发、应用开发
数据分析应用	自助取数	EasyFetch	自助取数工具，基于指标、维度查询数据	运营、产品、分析师
	网易有数	EasyBI	可视化数据分析报告制作平台	分析师、运营
	网易大屏	EasyScreen	可视化大屏	分析师
	数据填报	Easyfill	数据填报	运营
	移动端	EasyMobile	移动端报表	运营、分析师
	复杂报表	EasyReporter	多表头复杂式报表设计	分析师
	数据门户	EasyDPStudio	基于有数报表实现一站式建站，满足个性化数据门户建设	数据开发
	智能预警	Easywarning	基于报表数据的智能预警	分析师、运营
标签应用	标签工厂	EasyTag	一站式标签管理系统，包括标签的加工，ID-Mapping，人群的圈选、放大等功能	运营、数据开发

EasyData 数据中台支撑技术工具列表

全链路打通，形成产品闭环。比如任务运维中心与有数报告打通，可以看到一个任务影响了哪些下游报表。

组件式产品架构，允许业务根据场景搭配产品使用，形成面向不同场景的解决方案能力，例如数据质量解决方案、成本优化解决方案。

轻型易用、降低用户门槛，尤其注重非技术人员的交互体验。

我们研发这些工具产品，投入了 20 多个系统研发人力，因为公技可以把这些工具产品复用给其他的业务，所以对于单个业务来说，成本还是可以接受的。

第四，数据产品构建。

最后，就是业务支撑。我们通过构建数据产品，帮助业务达成业绩目标。我们的重点是商品和供应链。分别研发了商品运营系统和供应链辅助决策系统，大幅度提升了业务决策的效率。数据产品团队，我们有 10 个人的团队规模，主要负责数据产品研发。

总结数据中台项目成果

经过耗时 1 年半，实际执行一年的时间，我们完成了电商数据中台的搭建，并产出了一些阶段性的成果，对于成果这部分，你可以重点参考一下，因为你也可以通过这种方式，说服你的老板启动数据中台的建设。



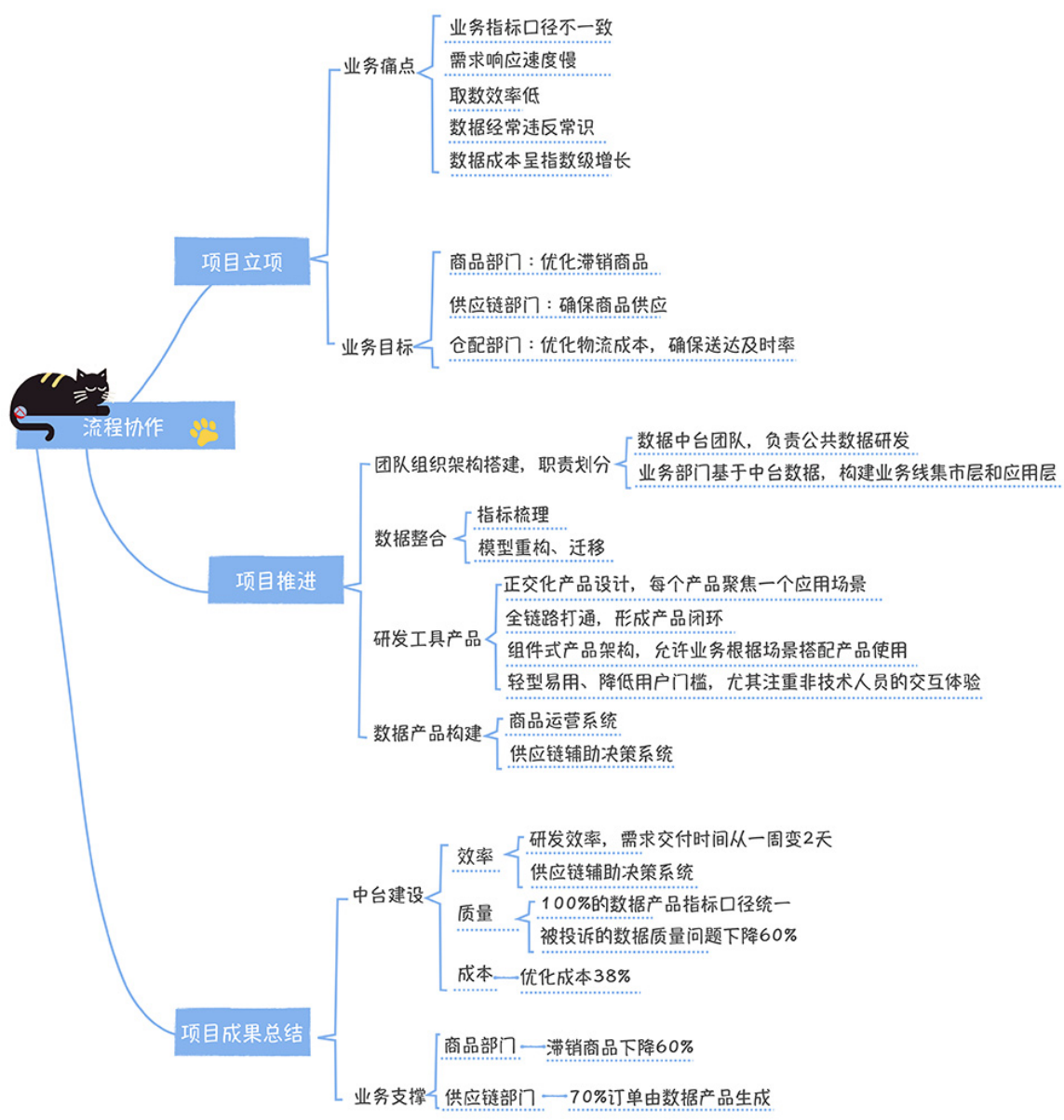
业务满意度大幅提升！从3.8提升到4.2

网易电商数据中台成果汇报

数据产品深入到业务运营，商品运营实现了滞销商品下降 60%，大幅提高了库存周转，降低了业务的成本。供应链辅助决策系统，每周有 70% 的订单由系统自动生成，推动给采购系统。用时一年时间，我们基本达成了在 2018 年底我们设定的中台建设目标。

课堂小结

这节课，我用网易电商建设数据中台的案例，从项目立项、推进、成果总结，让你看到了数据中台在企业落地的完整过程。最后，我想用一句话来总结今天的内容，那就是：数据中台从业务中来，到业务中去！同时我还想送给要建设数据中台的人一句话，“建设数据中台，不仅需要技术的思维，更需要管理者的视角”。



思考时间

学完最后一节课，你可以看到，所有关于数据中台的建设规范，我们最终都沉淀到了工具产品中，形成了产品化的解决方案，你知道这是为什么么？

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。

课程预告

5月-6月课表抢先看

充 ¥500 得 ¥580

赠 「¥ 99 运动水杯+ ¥129 防紫外线伞」



【点击】 图片，立即查看 >>>

© 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 (二) | 数据被加工后，你还要学会使用和管理数据

下一篇 结课测试 | 建设数据中台的这些知识，你都掌握了吗？

精选留言 (4)

写留言



吴科

2020-05-06

数据中台建设，我们踩过的坑

- 1 要得到高层的支持，由公司层面推进。最好由CEO直接负责。
- 2 统一数据口径，数据中台是唯一的数据输出平台。业务部门不用也得用了。
- 3 与业务部门利益绑定okr对齐，让业务参与数据中台的建设。

4 其他技术部门的联动，毕竟数据部门是下游系统，凡事都要靠上游输送数据。...

展开 ∨

作者回复: 说的挺好的，总结的很到位！

◀



👍 6



leslie

2020-05-06

CICD、DevOps、SRE其实整体上是都是相会补充的东西；中台其实就是这个的一个合理的展现。记得曾经和赵成老师和石雪峰老师都聊过相关的东西。

工具和平台是相互补充的，中台其实其实是一个平台，平台上面有各种各样的工具；工具提升的是各种事情的效率，平台是整合在一起让他们产生关联；没有关联数据其实是没有价值的数据。这就像有些东西我们为何要删除或者下架-它的存在没有价值了。...

展开 ∨

作者回复: 说的挺好的，尤其是中台体现的是效率这部分。

但是有一点，我们的认识上有一些偏差，就是中台和平台，还不能混在一起。中台，不是工具，工具能够支撑中台的构建，但是工具本身并不是中台。中台你可以把某个数据服务，称为中台，中台一定是有数据，带业务属性的。而工具本身，或者由工具组成的平台，并不具备业务属性，这是中台和平台最本质的区别。

感谢你的评论，我觉得除了上面我挑的这个瑕疵，其他的理解都超棒，尤其是关于CI/CD，Devops，SRE这部分，其实不管是技术中台，还是数据中台，很多原理都是相通的，这对于后面你去搞AI中台，还是其他什么中台，都是有借鉴意义的。

◀



👍 1



朱时

2020-05-08

你们的这个中台系统没有算法相关的内容，现在数据中台有得还会包括推荐服务、CV服务、语音服务、NLP服务等，这些如果要纳入中台算是前台服务还是中台服务呢

作者回复: 你好，数据中台并不包括语音服务，NLP服务，算法可以使用数据中台提供的数据，但是不是说数据中台直接提供AI能力。

这些如果要纳入中台，我觉得也是AI中台，不是数据中台。我觉得语音服务、NLP服务，这些可以作为中台，因为他们前面还有业务应用。

◀



aof

2020-05-07

好快啊，不知不觉已经最后一课了！

这篇专栏是我买的将近20个专栏里面最贴近我工作内容的专栏了，虽然专栏里面老师讲的一些东西我还没有在工作中实践过，但是读了这么详细的数据中台教程之后，在将来的实践过程中也不慌了，心里也有底了。...

展开 ∨

作者回复: 感谢@aof 一路相伴~

对数据中台的探索还在不断深入推进，希望后续有更多的交流机会~

