参考：<http://www.cyzone.cn/a/20160313/291400.html>

**大道至简的数据体系构建方法论：两步就让你打造出数据化运营的核心支柱！**

永洪科技副总裁王桐2016-03-13

标签: [数据体系构建](http://www.cyzone.cn/tag/%E6%95%B0%E6%8D%AE%E4%BD%93%E7%B3%BB%E6%9E%84%E5%BB%BA/)[数据分析](http://www.cyzone.cn/tag/%E6%95%B0%E6%8D%AE%E5%88%86%E6%9E%90/)



很多企业已经意识到，一个系统化的数据体系将是数据化运营的核心支柱。那么，企业该如何清晰地打造自己的数据体系呢？作者将根据多年经验总结用简朴的语言告诉读者一套大道至简的方法论。

本文是“数据化运营方法论系列”文章的第二篇。第一篇[《大道至简的数据分析方法论》](http://yonghongtech.com/html/news/company/2016/0129/460.html)之后的讲的是“不知道该怎么分析”的问题，本文讲的是“不知道该分析什么”的问题。第一篇文章更微观，站在个人分析师角度，本文更宏观，站在公司层面进行讲解。

与“不知道该怎么分析”一样，“不知道该分析什么”同样是很多人常问的问题之一。事实上，如果知道了方法，虽然不能做到没有一蹴而就，但是也能明晰如何一步步坚实地打造属于自己的数据体系路径。

与第一篇文章一样，本文会用最简单质朴的语言来讲清楚数据体系构建的路径。简单来讲，就是先梳理出数据指标体系，再将其落地到BI（商业智能，其实叫业务智能更对味）系统里。

**一，由上至下地梳理数据指标体系**

**1.确定目标**

这是第一个应该问自己的问题。花大力气做数据分析，最终为了什么呢？如果这都没想清楚，那数据体系肯定无从下手。

是想提高用户活跃度、增加用户、增加销量，还是别的什么目标？这么一想，好像我都想要。都想要没有问题，但是会让工作的边界无限蔓延，导致事情无法推进。所以，应该从最关心的那个目标/KPI入手。

那么，什么问题才是我们最需要关心的目标呢？

对于不同领域、不同阶段的公司和不同角色的用户而言，这个问题的答案都不一样：对于很多公司老板来说，利润就是他们最关心的目标；对于非售卖产品/服务的公司或政府而言，也许客户满意度是最关心的目标；对于交易平台类公司或早期电商公司而言，利润不是重点，交易量是最关心的目标。

最关心的目标搞定了，下面是不是可以解决都想要的问题了呢？并不是这样。大数据带来的最大一个误区就是数据量和字段数越多越好。但是，在真正解决具体业务问题时，我们一定是从大数据的全集中切出相关的一个子集来使用的。

对于单人而言，无论是老板还是执行层，同时关注的目标/KPI都不宜过多。同时看几十个KPI，想象一下也知道会很晕，且耗费时间。但是，对企业而言确实有很多KPI都是非常重要的。这该怎么办？可以分解到多人，即不同角色一起协作，每个角色关注自己的目标，所有角色合在一起是公司所有目标/KPI的全集。

假设老板最关注的目标是利润，利润=收入-成本，可以将这个目标分解为由销售总监来关注收入，运营总监来关注成本。当然，并不是说老板不能看收入，而是把常规性的关注目标锁定在一个可行的范围之内。

**2.分解指标**

目标确定了，下一步是分解出相关的指标。

针对目标，需要哪些指标来监控或分析能达成目标呢？比如利润，相关指标就是收入和成本，当然这太粗了，收入有哪几类，成本有哪几类，都应该考虑进去。比如对于零售行业的销售额，可以分解为客流量、进店率、购买率、客单价和复购率等。

所以，分解的方式有很多种，需要遵循MECE原则（完全穷举，相互独立）。

**3.细化字段**

针对指标的计算公式，涉及到哪些字段，分别在哪些库的哪些表里，是否需要数据清洗，清洗规则是什么等。

比如购买率，是通过公式“购买人数/进店人数”算出来的，购买人数又是对“客户ID”进行计数计算得出来的，这些指标涉及到的字段对应到数据库里哪张表的哪个字段，需要梳理清楚，这部分就需要IT人员或数据库管理员的介入和配合了。

**4.非功能需求**

上述第3步完成之后，我们其实已经算是梳理完了指标体系，可以落地了，但为了让最终形成的数据系统更加完备、友好、可用，还需要一些非功能需求的梳理。

UI：偏好什么样的展示风格，这点看着无关紧要，但实际上用户每天都会与数据系统打交道，美观、体验好的系统UI会让用户更加喜欢。

页面流：哪些相关指标摆放到同一个报告页面上，页面之间的层次关系如何，用户可以在页面之间如何跳转。

权限：谁能看哪些数据范围，谁能看哪些字段和指标，需要有统一的权限控制，避免出现数据安全问题。

ETL：数据从数据源同步到分析系统的频率如何，规则如何。

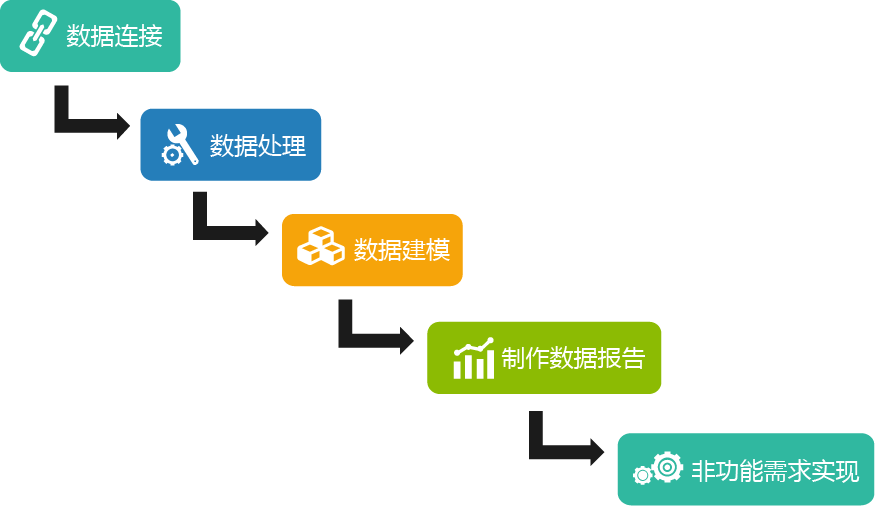
集成：是否需要在界面、预警消息等层面与其它系统进行集成。

性能：看不见摸不着，但是直接决定系统可用性。如果数据量大时需要几分钟甚至几十分钟才能看到结果，相信这个系统就不会有人愿意用了。

**5.系统实施**

上述4项完成之后，我们就形成了《数据运营系统需求文档/实施方案》，即可落地到数据运营系统里，然后，再根据报告页面数量、数据准备复杂度等确定工作量和时间计划。

**二．由下至上地实施落地到BI系统**



**1.连接数据**

根据需求文档/实施方案，一步步进行系统搭建工作。这个系统有的企业称之为大数据平台，有的企业称之为BI系统。大数据平台的范畴会更广一些，但对企业数据化运营而言，BI一定是核心构成。

那么，无论是开发还是基于像永洪科技一样的第三方工具快速实施，系统搭建的第一步都是连接各个数据源，打通和各个数据源之间的通路。

在企业里，数据环境往往是异构的，数据源可能包括数据库、Hadoop系列平台、Excel文件、日志文件、NoSQL数据库、第三方接口等，需要对每种数据源都有快速友好的对接方式。

最终，我们在系统里能看到所需要的各个数据源中所有的表格和字段。

**2.数据处理**

数据源里的数据往往是有或多或少的不规范性存在的，比如有重复记录，比如有遗漏的空值，比如有明显不合理的异常值（比如有2020年的成交订单），还可能有同一个事物在系统中存在多个名称的情况。

这些数据如果不做一些处理或称之为清洗的工作，是会对分析的准确性产生很大影响的，所以需要做些预处理。这个过程往往是最耗时、最枯燥的，但也是十分重要的。

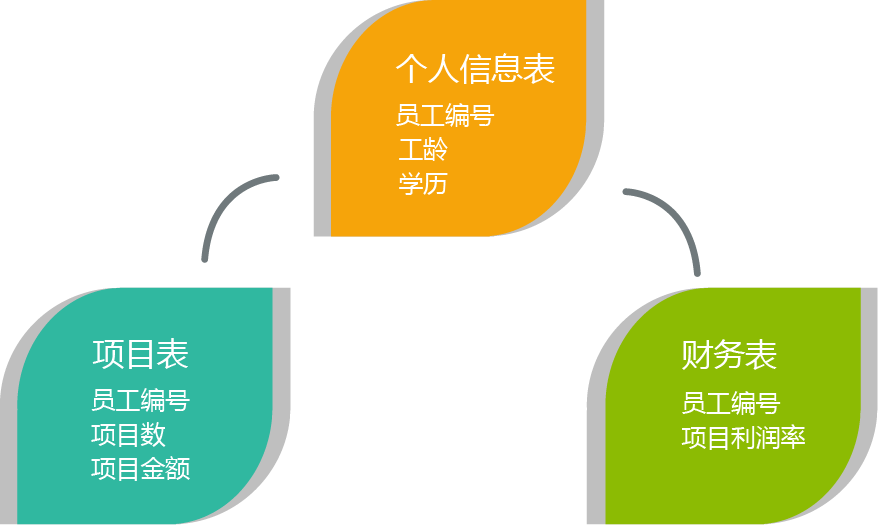
作者提醒：这个环节的问题将在下一篇《大道至简的数据治理方法论》文章中再深入探讨。

**3.数据建模**

数据处理好了，下一步就该做数据建模了。

一提到建模，非技术背景的用户就生畏，觉得高深不可理解。其实建出的模是个什么东西呢？简单来讲，把多张表关联到一起，就是一个数据模型。

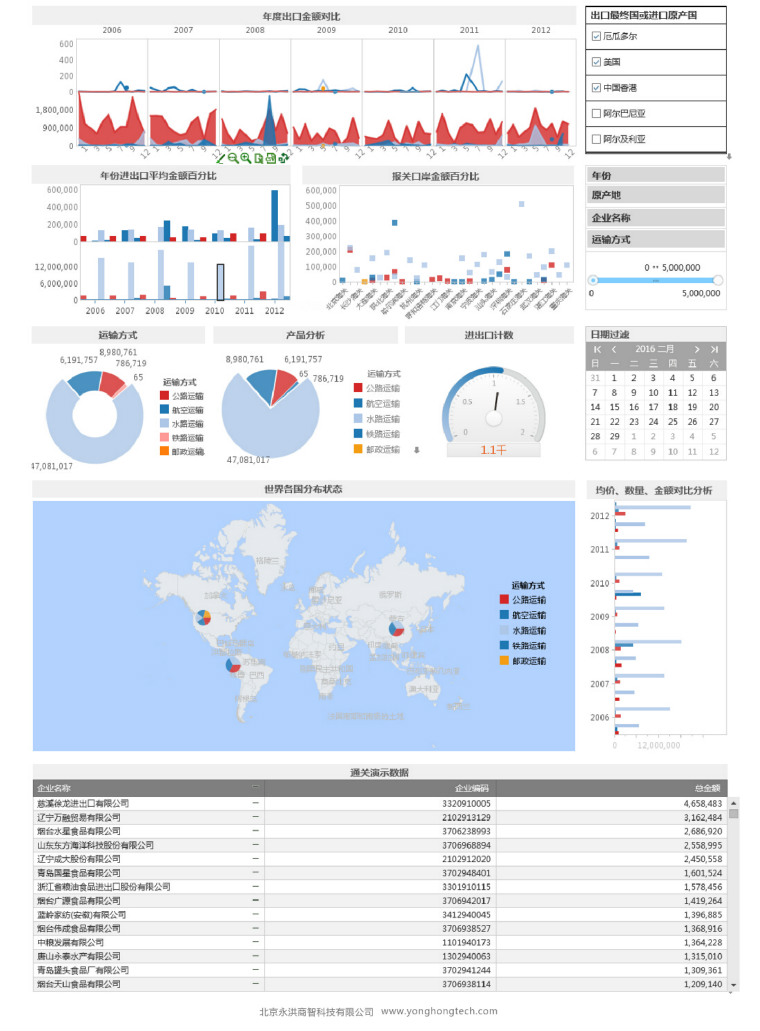
比如，公司要做绩效分析，需要员工的工龄、学历、项目数、项目金额、项目利润率等指标，其中工龄、学历在个人信息表里，项目数、项目金额在项目表里，项目利润率在财务表里，这三张表有个共同字段“员工编号”，通过这个字段把这三张表关联起来，这就是一个数据模型，一个绩效分析主题的数据模型。



**4.制作数据报告**

基于建好的数据模型，我们就可以开始制作数据报告了。

数据模型提供了基础数据和字段，按照需求将它们以公式进行组合，用合适的图表类型进行展示，将相关指标摆放到同一个报告页面上，配置好页面之间的层次关系和跳转关系。以下是基于永洪科技一站式大数据分析平台制作的Demo。



**5.非功能需求实现**

经过第4步之后，我们的数据系统已基本成型，剩下的就是实现上述的各个非功能需求了。这样，一个完备、友好、可用的数据运营系统就上线了。

上线并不是工作的终点，业务需求时刻都会变化或新增，需要能够快速迭代调整，数据处理、建模、制作数据报告等操作需要高度工具化，以保证灵活可配置。第三方工具对比自开发的优势也在这点上体现尤为明显。

归根结底，做数据的目的要么是为了提升管理（节流），要么是业务创新（开源）。一个系统化的数据体系将是数据化运营的核心支柱。

编者按：由于时间关系和水平有限，文中或有不妥之处还请读者多多谅解。欢迎读者通过tylerwang@yonghongtech.com或微信号tyler\_wangtong与本文作者永洪科技副总裁王桐取得联络，一起探讨数据分析方法论和数据体系构建方法论等相关话题。