

指标管理体系基础知识及自助分析平台的使用

一、引子



指标定义整体概览

实际使用过程中，一般就使用派生指标和复合指标
这个图现在看不懂没关系，大家稍微看一下，有个概念。



上图看不懂？没关系，我们先来几个常见指标来热热身：

part1:

1. MAU：月活跃用户数量

a. 口径定义

i. 月

ii. 活跃

iii. 用户去重数 【count(distinct uuid)】

2. DAU：日活跃用户数量

a. 口径定义：

i. 日

ii. 活跃

iii. 用户去重数 【count(distinct uuid)】

3. 日充值用户数量

a. 口径定义

i. 日

ii. 充值

iii. 用户去重数 **【count(distinct uuid)】**

4. 日充值总金额

a. 口径定义

i. 日

ii. 充值

iii. 充值总金额 **【sum(amount)】** (单位)

5. 男性日充值总金额

a. 口径定义

i. 日

ii. 充值

iii. 充值总金额 **【sum(amount)】** (单位)

iv. 男性

6. 安卓男性日充值总金额

a. 口径定义

i. 日

ii. 充值

iii. 充值总金额 **【sum(amount)】** (单位)

iv. 男性，安卓



总结：

前面4个指标口径都是由3部分组成

a. 统计周期 (日, 月)

b. 事件

c. 聚合逻辑 (用户去重数 **【count(distinct uuid)】** , 充值总金额 **【sum(amount)】**)

第5,6个多了一些条件限制

5多了: 用户性别= '男性'

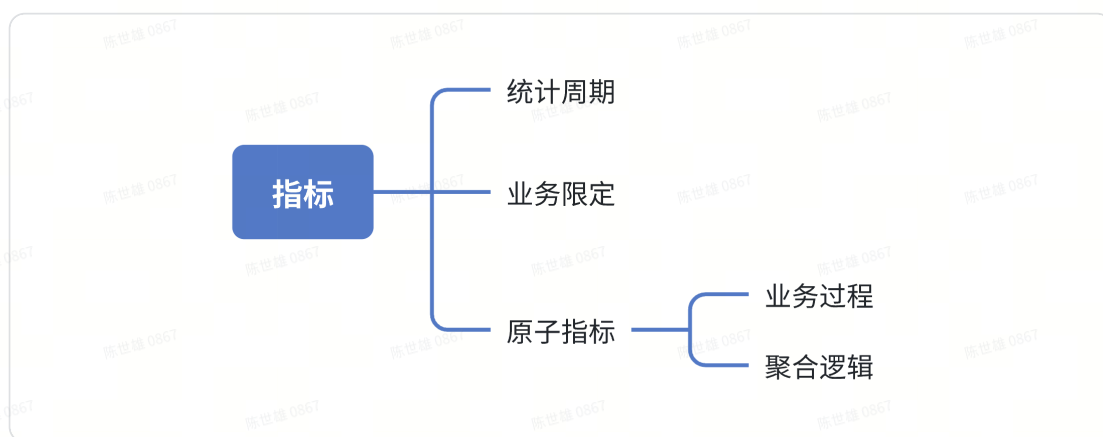
6多了: 用户性别= '男性' and 手机平台= '安卓'

还有：充值金额>100（业务限定）

所以：指标的组成：

- a. 统计周期
- b. 业务过程（事件）
- c. 聚合逻辑
- d. 业务限定

我们为了复用，把业务过程和聚合逻辑合并叫做原子指标，于是就如下图所示：



所以最终指标由3部分组成

- 1. 统计周期
- 2. 原子指标
- 3. 业务限定

part2:

我们再来看一下稍微复杂一点的指标：

- 1. 天级的聊天室进房率
- 2. 注册用户的次日留存率

这时我们发现：找不到业务过程（事件）了，我们先回到它们的计算公式：

- 1. 聊天室进房率=聊天室的进房次数/聊天室的曝光次数 * 100%
 - a. 聊天室的曝光次数
 - i. 业务过程：曝光聊天室
 - b. 聊天室的进房次数
 - i. 业务过程：进入聊天室

2. 新用户的次日留存率 = $\frac{T\text{日活跃且}T-1\text{日新注册的用户数}}{T-1\text{日新注册的用户数}}$

a. T日活跃且T-1日新注册的用户数

i. 业务过程：活跃

ii. 业务限定：用户的注册时间是T-1

b. T-1日新注册的用户数

i. 业务过程：注册

可以看出，分子和分母都有业务过程，进一步的分子和分母都可以组合成一个指标；

也就是说，这2个指标都可以由2个指标通过除法计算而来。



所以我们对指标进行分类：

1. 原子指标：由业务过程和聚合逻辑组合而成的；单独使用没有意义；

2. 派生指标：通过原子指标+统计周期+业务限定派生而来的，例如：part1的那些指标

3. 复合指标：通过派生指标运算（加减乘除）而得到的，例如：part2的那些指标

二、基础概念：

1. 事件模型：

a. 5W2H：Why、When、Who、Where、What、How、【How much】

b. 题外话：埋点设计的时候也可以这么思考，补全埋点属性

2. 维度：

a. 5W + How

3. 度量：

a. How much

4. 业务过程

a. 粒度声明

i. 明细，小时，天

ii. 公会，用户

iii. 支付方式



1. 可以简单理解为：事件 （事件是粒度最细的一种业务过程）
2. 举个非最细粒度的例子：用户每天的充值情况 （这里的粒度就是用户，按天【唯一key就是用户id&充值日期】，是由明细业务过程（用户充值）按天汇总而成的一个业务过程）

b. 业务过程定义说明

- i. 过程（充值，支付，曝光，点击）
- ii. 结果（剩余趣豆数,库存量）无过程【初学超纲，可以不管】

5. 原子指标

a. 业务过程

b. 聚合逻辑

- i. 度量值：聚合算子括号内的那个字段
- ii. 聚合算子：sum，count，count distinct

6. 派生指标

a. 统计周期

- i. 小时级，天级，周，月，年

b. 原子指标

- i. 活跃用户数
- ii. 支付金额

c. 业务限定

- i. 维度+操作符+比较值
 1. 例如：用户性别 = ‘女’
 2. 城市 包含（‘厦门’，‘深圳’，‘广州’）

7. 复合指标

a. 派生指标

b. 运算：加、减、乘、除

- i. 例如：各种转化率：曝光点击转化率=（点击次数/曝光次数）* 100%
- ii. 各种留存率：新增留存率，活跃留存率

8. 统计粒度

- a. 看指标的角度
- b. 维度
 - i. 例如：按平台统计

三、举例子：

💡 派生指标：按平台统计天级女用户的注册成功次数

指标组成，SQL和自助分析对比

	定义	SQL	自助事件分析截图
原子指标	注册成功次数	<pre>SELECT COUNT(1) 度量值 FROM register_success WHERE event_time = '2024-06-06 00:00:00' GROUP BY platform</pre>	
统计周期	天级	<pre>SELECT COUNT(1) 度量值 FROM register_success WHERE event_time BETWEEN '2024-06-06 00:00:00' AND '2024-06-10 00:00:00' GROUP BY platform</pre>	
业务限定	用户性别=‘女’ 事件时间：6.6~6.10	<pre>SELECT COUNT(1) 度量值 FROM register_success WHERE event_time BETWEEN '2024-06-06 00:00:00' AND '2024-06-10 00:00:00' AND gender = '女' GROUP BY platform</pre>	
统计粒度	看数的角度 按平台	<pre>SELECT COUNT(1) 度量值 FROM register_success WHERE event_time BETWEEN '2024-06-06 00:00:00' AND '2024-06-10 00:00:00' AND gender = '女' GROUP BY platform</pre>	
完整的	6.6~6.10期间按平台统计天级女用户的注册成功次数	<pre>SELECT COUNT(1) 度量值 FROM register_success WHERE event_time BETWEEN '2024-06-06 00:00:00' AND '2024-06-10 00:00:00' AND gender = '女' GROUP BY platform</pre>	

再回过头来看看开篇的那个图，再理解一下。

四、如何在自助分析平台上查询数据指标？

自助分析平台_事件分析 国天玑引擎使用说明（含权限申请） 国高级分析-事件分析

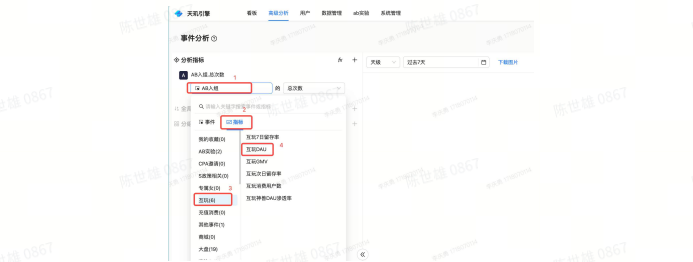
 **自助查询**

- 1. 原子指标（难点）
 - a. 事件（难点） 国他趣埋点事件汇总文档
 - b. 聚合逻辑
- 2. 业务限定
- 3. 统计周期
- 4. 分组条件



 **指标功能**

- 1. 选择指标（把自助查询的原子指标做了一层封装，降低使用门槛）
- 2. 业务限定
- 3. 统计周期
- 4. 分组条件（统计粒度）



进一步：复合指标：



五、参考资料

大数据之路：阿里巴巴大数据实践.pdf 第158页~第164页 （9.2.2 指标体系）

华为数据之道pdf高清版.pdf 第215页~第219页 （5.3.5 指标设计）

这里面讲的也都比较抽象，不太好理解，大家可以结合上面的一些例子（特别是SQL）去理解。

数据开发需要更进一步： 大数据之路：阿里巴巴大数据实践.pdf

a. 维度设计 （第10章）

b. 事实表设计 （第11章）

c. 数据质量 （第15章） & 华为数据之道 （第8章） 【这部分华为感觉更好】