

数据分析

课程目标

学习完本课程后，你将能够：

1. 了解指标的定义、价值和评价标准
2. 掌握指标体系的构建方法
3. 掌握阿里云大数据产品MaxCompute、DataWorks 的基本应用

课程目录

1. 数据分析指标与实现

1.1 什么是数据指标

1.2 优秀的数据指标的特征

1.3 数据分析指标的实现

2. 基于阿里云平台进行数据分析

什么是指标、数据指标

“如果你不能衡量，那么你就不能有效增长”

——现代管理学之父彼得·德鲁克

对企业或组织来讲，衡量即按统一标准来定义、评价业务情况，
这个标准即**指标(Indicator)**。



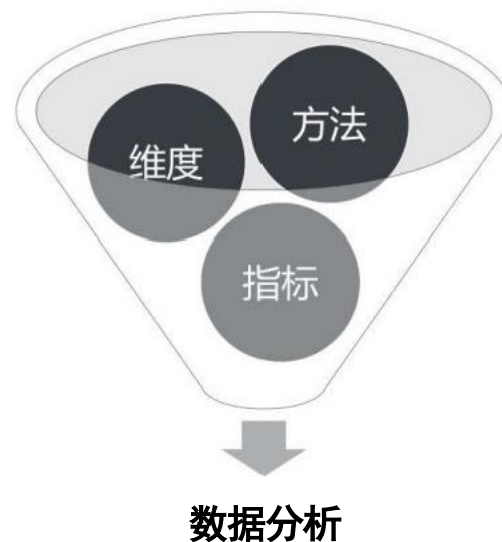
指标就是说明总体数量特征的概念及其数值的综合。我们在数据库、数据仓库中进行数据操作，包括统计、查询、分析等，其针对的数据字段(这些字段与产品、业务相关，对产品和业务有参考价值)，本质上都是**数据指标**。

企业指标体系则是若干个反映企业业务、企业生产运营、企业产品等数量特征的相对独立又相互联系的数据指标所组成的有机整体。

数据分析与数据指标

数据分析是指用适当的统计分析方法对收集来的大量数据进行分析，将它们加以汇总和理解并消化，以求最大化地开发数据的价值，发挥数据的作用。一句话，数据分析是为了提取有用信息和形成结论而对数据加以详细研究和概括总结的过程；数据分析离不开数据指标，数据指标贯穿数据分析过程。

- ＞ 数据分析是一个过程：确定分析目标-->确定数据指标-->获取数据-->生成数据指标-->分析数据指标-->得出结论；
- ＞ 数据指标是数据分析的载体，是通过数据指标实现数据分析的目标；
- ＞ 各种维度的数据指标即为数据分析的结果展现形式；
- ＞ 数据指标结合分析方法得出新的指标，通过数据指标获得分析结论。



指标的组成

完整的数据统计指标：

- ＞指标的名称，说明所反映现象数量特征的性质和内容
- ＞统计的时间界限和空间范围
- ＞计算方法（修饰词）

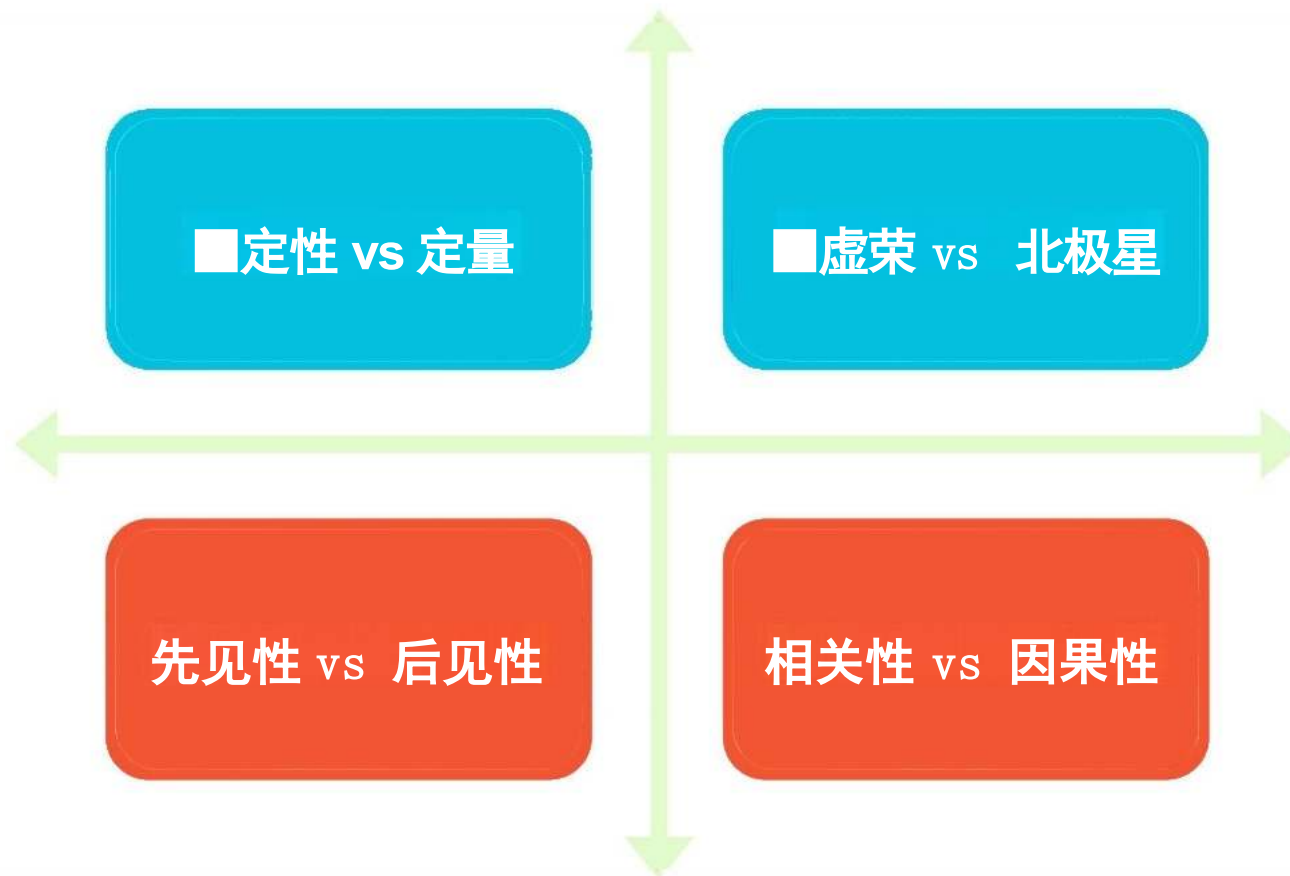
例如

截止2020年1月1日A学校在岗教职工人数

618这一天B店铺通过直播带来的利润

指标	
客户	活跃数
	新增数
财务	销售额
	销售利润

数据指标的分类



定性和定量

定性指标

非结构化的、经验性的、揭示性的、难以归类的指标

定量指标

可以被计数和衡量的指标

【例】在绩效考核当中

- 管理层：适宜采用定量成分较多、约束力较强、独立性较高、以最终结果为导向的考核指标
- 普通员工：适宜采用定性成分多、需要上下级随时充分沟通、主要以工作过程为导向的考核指标

虚荣和北极星

虚荣指标

容易给人留下印象，但是无法用于决策的指标

北极星指标

可以指明工作的方向，帮助改进商业模式，决策下一步行动的指标

【例】

■ 虚荣指标：总注册用户数

■ 北极星指标：总活跃用户数、总活跃用户占比

先见性和后见性

先见性指标

用于预测企业未来情况的指标
通过对未来的预测，制定利益最大化的策略

后见性指标

用于揭示当前存在问题的指标
发现存在的问题后通过干预，减少损失

【例】

- 先见性指标：下半年每个月的预测销量
- 后见性指标：用户流失率变化

相关和因果

相关指标

一个或多个指标变化的同时，另一个或多个指标也会随之变化
但是指标之间不存在先后的时间关系

因果指标

一个或多个指标的改变，能够对另一个或多个指标产生某种作用
指标之间的变化存在先后的时间关系

【例】

■ 相关指标：雪糕销量和溺亡人数

■ 因果指标：商品单价和销售额

数据指标的其他分类方式

不同行业、不同企业、同一企业不同部门内指标分类不近相同，常见数据指标的分类：

- ＞按通用性分：宏观指标、特定指标，如互联网行业常见的宏观指标PV、UV，具体到某业务的销售情况则为特定指标；
- ＞按重要性：KPI指标、普通指标，即核心关键指标(按部门、按业务、按时间等)、一般指标；
- ＞按指标内容或其数值表现形式：总量指标、平均指标、相对指标，如上半年总收入、上半年月均收入、上半年增长率
- ＞按管理功能：描述指标、考核指标、监控指标等，
- ＞其他分类

常用的数据指标

不是所有的数据指标都叫指标，只有对当前业务有参考价值的指标才可称作指标。

对于互联网电商而言，常见指标如下：

网页基础指标：

- **PV（页面浏览量）** 用户每1次对网站中的每个网页访问（成功访问/进入）均被记录1次。用户对同一页面的多次浏览，浏览量累计。在一定统计周期内用户每次刷新网页1次也被计算1次；
- **UV（独立访客人数）** 访问网站的一台电脑客户端为一个访客。统计周期内相同的客户端只被计算一次；
- **VV（用户访问次数）** 当用户完成浏览并退出所有页面就算完成了一次访问，再次打开浏览时，VV数+1；
- **跳出率** 表示用户来到该页面后，没有进行任何操作就直接离开的比例，代表了对客户的吸引力；
- **退出率** 用户从某个页面进入网站，最终从该页面退出的比例；
- **平均访问时长** 指在特定统计时间段内，浏览网站的一个页面或整个网站时，用户所停留的总时间除以该页面或整个网站的访问次数的比例；
- **转化率** 在一个统计周期内，完成转化行为的次数占推广信息总点击次数的比率；
- **回购率** 指用户对商品或者服务的重复购买次数（回头客）。

常用的数据指标

不是所有的数据指标都叫指标，只有对当前业务有参考价值的指标才可称作指标。

对于互联网电商而言，常见指标如下：

用户基础指标：

- ＞ ARPU 每个用户平均收入， $ARPU = \text{总收入} / \text{用户数}$ ；
- ＞ **新增用户** 新注册使用电商系统的用户，又通常分为：**日新增(DNU)**、**周新增(WAU)**、**月新增(MAU)**，注意重新注册用户不算，这是衡量营销推广渠道效果的最基础指标；
- ＞ **活跃用户** 在特定的统计周期内，成功启动过、使用过产品的用户；活跃用户数根据不同统计周期可以分为**日活跃数(DAU)**、**周活跃数(WAU)**、**月活跃数(MAU)**；注意同周期内用户个数不累计；
- ＞ **留存率** 在某一统计时段内的新增用户数中再经过一段时间后仍启动该应用的用户比例($\text{留存率} = \text{留存用户} / \text{新增用户} * 100\%$)。通常重点关注次日、3日、7日、30日即可，并观察留存率的衰减程度；留存率又分为：**次日留存率**、**7日(周)留存率**、**30日(月)留存率**等等；

常用的数据指标

不是所有的数据指标都叫指标，只有对当前业务有参考价值的指标才可称作指标。

对于互联网电商而言，常见指标如下：

用户基础指标：

- **流失率**：指那些曾经使用过产品或服务，由于各种原因不再使用产品或服务的用户，**用户流失率=某段时间内不再启动/使用应用的用户/某段时间内总计的用户量**，流失率和留存率有紧密关联，流失率高既留存率低，但活跃度不一定高，因此需要综合分析，也可以按周期统计流失率；
- **一次性用户**：即新增日后再也没有启动过/使用过应用的用户；在实际应用中一般认为至少超过7天时间才能够定义是否是一次性用户，即一周以上时间未使用/未登录应用。
- **使用时长**：统计时间段内，某个设备从启动应用到结束使用的总计时长；
- **启动/登录次数**：统计时间段内，用户打开/使用应用的次数；
- **使用间隔**：用户上次使用应用的时间与再次使用的时间差。

数据指标的价值

数据指标的价值：

- 改变行为
- 促成目标：优化产品、优化定价、优化市场定位等

【例】指标显示将“立刻注册”改成“免费注册”可以提升20%的转化率，那么就应该将文案优化

【反例】销售员的季度奖金应该与手头上的订单数这一指标挂钩吗？

销售团队会为了个人收入制造大量低质量的潜在客户，而不是想办法获得更多高质量客户。

应该与已完成的订单数量或订单的利润率挂钩。

课程目录

1. 数据分析指标与实现

1.1 什么是数据指标

1.2 优秀的数据指标的特征

1.3 数据分析指标的实现

2. 基于阿里云平台进行数据分析

优秀的数据指标的特征

可比较

- 不同的时间段
- 不同用户群体
- 不同竞争产品

本周转化率比上周高2%
本周转化率为5%

可理解

- 简单易懂
- 避免误导
- 业务解释性

Sales
销量?销售额?

是比率

- 可操作性强
- 可以比较
- 反映关联性

里程 vs 速度
当前速度 vs 区间平均速度

优秀的数据指标的特征

某APP公司面临一个选择：该在免费版中提供尽量丰富的功能以吸引新用户，还是该将这些功能保留在收费版本中，以促使用户为高级功能付费？

- 推出功能丰富的免费版不利于销售额的增长
- 而免费版功能过简又不利于新用户的增加。

方案一：

免费功能 A BC **D E**

收费功能 FG HIJ

日均新用户数 **100**

日均收入 **1000**

指数 = 10%

方案二：

免费功能 ABC

收费功能 **D EFG** HIJ

日均新用户数 **60**

日均收入 **1200**

指数 = 5%

指数 = 新用户数 / 收入

- 一个结合了二者的数据指标来帮助理解产品的改动对公司的整体业绩会有怎样的影响。
- 否则，可能会片面地为增加销售额而牺牲新用户的增长。

课程目录

1. 数据分析指标与实现

1.1 什么是数据指标

1.2 优秀的数据指标的特征

1.3 数据分析指标的实现

2. 基于阿里云平台进行数据分析

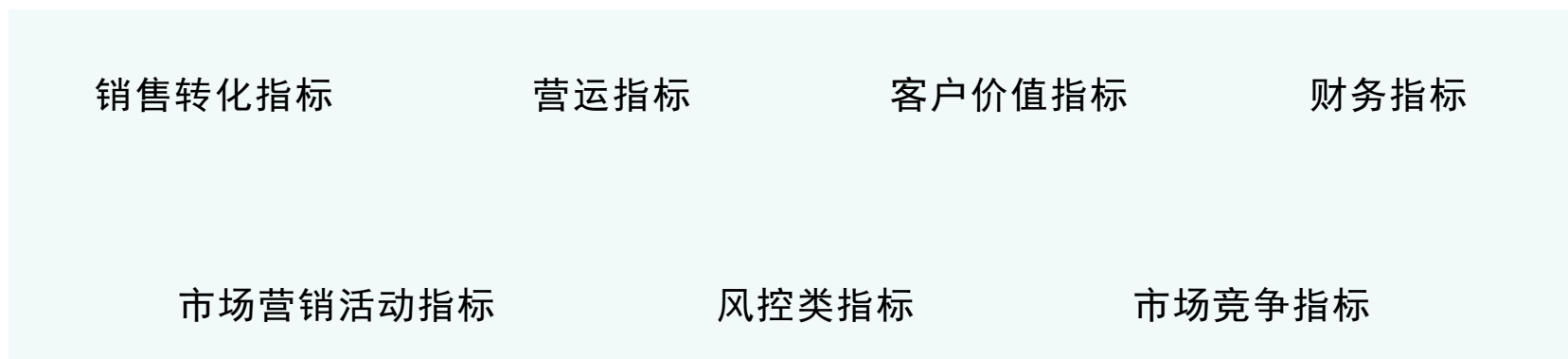
数据分析指标的实现

数据分析指标的实现过程也是从需求出发，首先明确分析的任务目标，结合企业业务，设计制定优秀的指标，基于分析目的，采取相应的分析方法，然后指标落地实施，最后考评指标分析结果与原需求目标是否匹配，是否能达到既定目标。



数据分析指标的设计维度示例

在企业指标体系中有许多指标来衡量企业的生产运行情况，衡量产品销售、用户质量、市场活动、风险控制的效果，基于不同的角度将这些指标进行维度划分，如下示例：



数据分析指标案例

某电商平台考察风险控制情况，业务主要是线上销售商品，客户购买商品，购买完成对购买过程、快递物流、购买商品进行线上评价，或者客户购买商品后由于各种原因线上投诉，还有就是进行线上退、换货等。

结合具体实际情况，设计风控类指标时，可以从三个方面考虑：买家评价、买家投诉、退换货等。

其中，买家评价指标可设计为：

指标名称	指标定义	指标作用
买家评价率	指某段时间参与评价的买家与该时间段买家数量的比值；	反映用户对评价的参与度，电商网站目前都在积极引导用户评价，以作为其他买家购物时的参考
买家好评率	指某段时间内好评的买家数量与该段时间买家数量的比值；	
买家差评率	指某段时间内差评的买家数量与该段时间买家数量的比值；	

数据指标的分析方法

数据指标通常分为定量和定性，定性指标即分类指标，如满意度高还是低、产品质量优还是差、一般；定量指标则是通过准确数量定义、精确衡量并能设定绩效目标的考核指标。

定性指标分析方法：

- ＞对比分析，如按时间分析每月满意度情况；
- ＞结构分析，如产品质量优、一般、差的占比分析；

定量指标分析方法：

- ＞分组分析，指标分组分析，企业按不同区域分组，考察某月营收情况；
- ＞结构分析，在统计分组的基础上进行分析；
- ＞交叉分析，即立体分析，指标在横向分析、纵向分析的基础上，从交叉、立体出发从低级到高级分析的综合分析方法

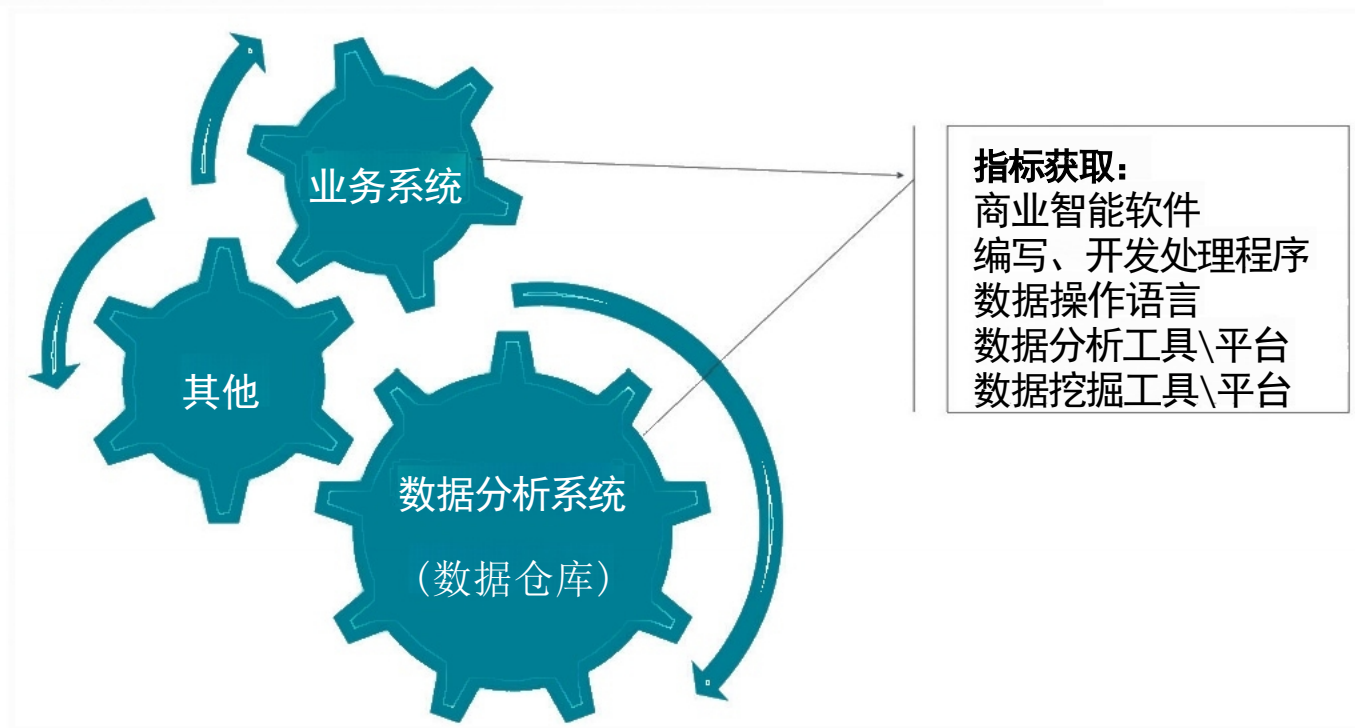
数据指标的分析方法

交叉分析案例：

假设A公司资产净收益率0.35%，营业利润率0.74%，每股收入为0.001元，而B公司资产净收益率10%，营业利润率12%，每股收入为0.3元，如果只看这些数据，横向比较明显B公司优于A公司，如果从纵向做对比，发现A公司一年来，收益率、利润率、每股收入都是逐月在增加，而B公司三项指标是逐月下降，对于投资者而言，往往认为A公司要比B公司更有吸引力。

数据分析指标的系统实现

指标定义完成后，就要考虑指标的系统实现了，即如何实现这些数据指标。理论上讲，数据指标可以从指标涉及数据的源系统(或其他非系统数据源)直接获得，也可以从专门的数据分析系统获得。

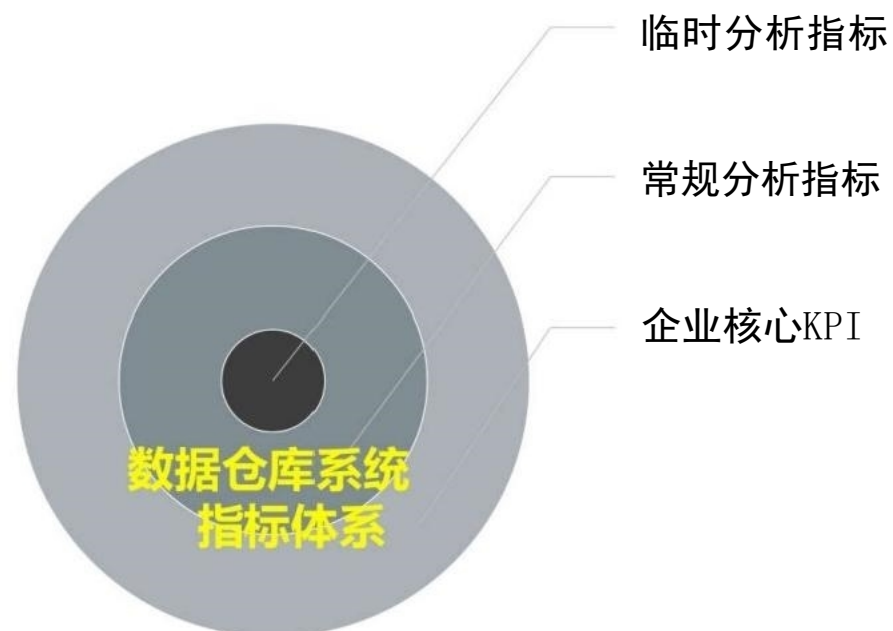


数据分析指标的系统实现

大数据、云计算和互联网发展到今天，数据量越来越大，维度越来越多。企业对数据、效率要求的逐步提高，也给大数据提供了展现能力的平台，基于大数据分析平台(数据仓库系统)获取各类数据分析指标成为企业数据指标分析的常见方式。

常见数据分析工具：

- >QlikView
- >Tableau
- >PowerBI
- MicroStrategy
- >FineBI
- >QuickBI



数据指标分析案例

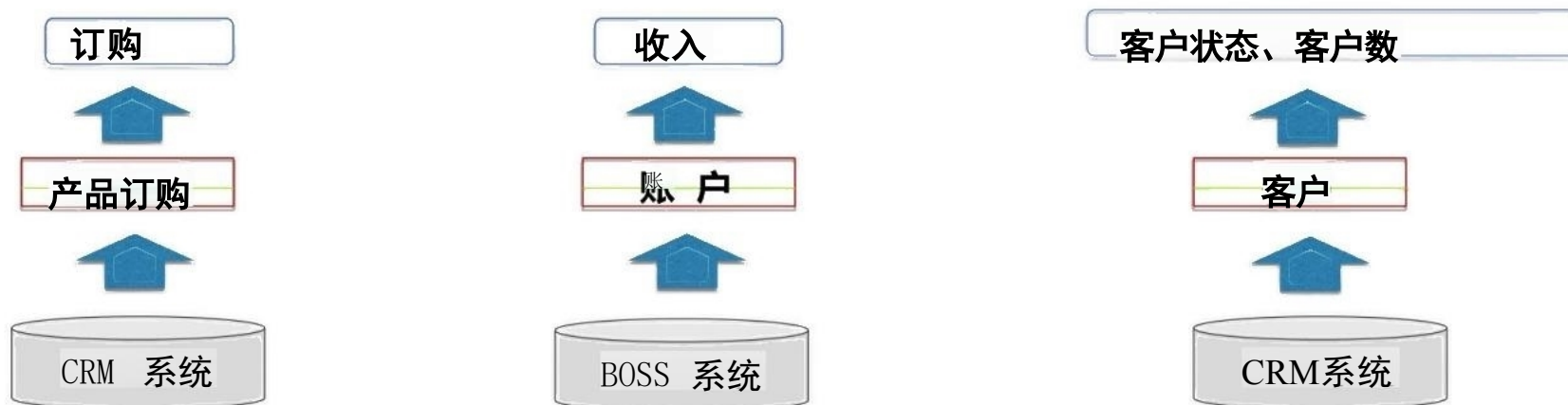
以某企业实现指标“**有效集团客户数**”为例，此指标的相关信息如下：

指标类别	指标名称	指标定义	单位	展现要素	
集客经营	有效集团客户数	截至统计周期末，订购了集团通信和信息化产品(不含校园V网、乡情网、家庭V网)，且近三个月累计通信和信息化收入大于0，状态在网的A、B、C、D集团客户数	家	本月值	同比变化 环比净增

说明：此指标涉及集团客户信息、集团客户订购产品信息、集团客户产品使用信息等，是一个综合性的指标，计算方法为“同比”和“环比”。

基于数据库实现数据指标分析案例

指标“**有效集团客户数**”的数据源涉及多个业务系统，其明细口径为“截至统计周期末，**订购**了集团通信和信息化产品（不含校园类产品、家庭类产品），且近三个月累计通信和信息化收入大于0，状态在网的集团**客户数**。”，数据来源分析如下：



说明：

CRM：承载客户管理、客户产品订购相关业务的系统；

BOSS：承载有关客户计费、账务相关业务的系统。

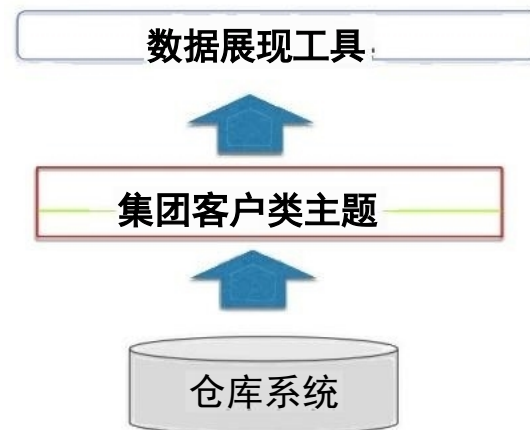
基于数据仓库实现数据指标分析案例

指标“**有效集团客户数**”的数据源涉及多个业务系统，其明细口径为“截至统计周期末，**订购**了集团通信和信息化产品（不含校园类产品、家庭类产品），且近三个月累计通信和信息化收入大于0，状态在网的集团**客户数**。”，在数据仓库中实现：

技术口径：

```
select count(distinct 客户ID)
From 客户表
Left outer join 订购表 on 客户表.客户ID= 订购表.客户ID
Left outer join 账户表 on 客户表.客户ID= 账户表.客户ID
Where 账户表.近三个月累计通信和信息化收入>0
And 订购表.订购产品名称 not in (校园类, 家庭类)
And 客户表.客户状态=“在网”；
```

更简单的情况是：客户表、订购表、账户表在数据仓库中有统一的集团客户统一视图



课程目录

1. 数据分析指标与实现
2. 基于阿里云平台进行数据分析
 - 2.1 MaxCompute基本操作**
 - 2.2 DataWorks基本操作

大数据分析平台/数据仓库

当今是互联网、云计算、大数据快速发展的时代，企业如何对海量、多源、多类型数据做有效的应用、分析，有效途径之一就是构建大数据分析平台，通过大数据平台来收集数据、存储数据、加工数据、分析数据、展现及应用数据。数据分析平台常见功能组成有：

数据采集

数据存储

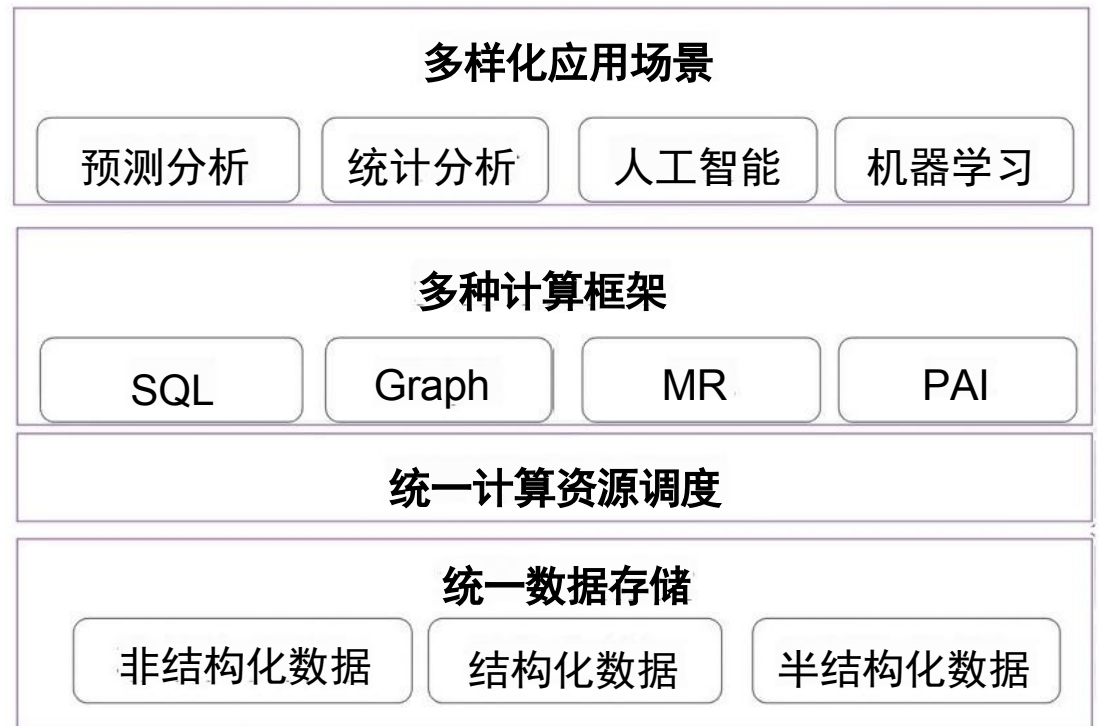
数据分析与处理

数据展现

全过程自动调度和智能运维

基于阿里云搭建数据分析平台/数据仓库

- 统一存储，减少不必要的冗余和跨库搬移数据
- 提供SQL/MR/机器学习/图计算等多种计算能力，避免为计算而搬数据
- 统一权限和计算资源调度，最大化的提升资源的利用率和弹性
- 统一元数据最大化数据统一管理



大数据计算服务-MaxCompute



MaxCompute

大数据计算服务MaxCompute, 由阿里云自主研发, 提供针对TB/PB级数据、**实时性要求不高的分布式处理能力**, 应用于**数据分析、挖掘、商业智能**等领域。阿里巴巴的数据业务都运行在MaxCompute。

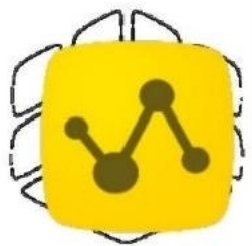
MaxCompute 是一种快速、完全托管的EB级数据仓库解决方案。

MaxCompute 的主要功能及作用

MaxCompute 致力于批量结构化数据的存储和计算，提供海量数据仓库的解决方案及分析建模服务。通常和 DataWorks 一起构建大数据分析平台。



MaxCompute 的使用流程



MaxCompute 所有的操作都是基于表，因此MaxCompute 的使用从安装配置环境后建表开始

MaxCompute



以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：
<https://d.book118.com/485223300013011201>