

一、企业数据分析能力简介

- 数据分析高度不可见，过程不可控，结果不可测量
- 数据分析要看人品，看 SQL 语句规不规范、取数准不准
- 概述第一部分

概述第一部分

企业数据分析能力简介

- ➔ ——企业决策的四个层次：战略、管理、运营、操作
- 企业数据分析能力的演进
- 企业运营和操作数据应用
- 数据管理基础知识与DMBOK知识体系

• 数据在企业中如何被使用

好的企业业务流程完全可视化 做事遵循流程，流程中存在决策点 数据分析要服务于决策，服务于流程中的决策节点

• 企业决策的四个层次



企业决策的四个层次：战略、管理、运营、操作 宏观：PEST 中观：企业上下游、SWOT、波特五力模型 微观：收入成本分析 战略：做什么？为什么做？ 战略分析用的都是外部数据，投资分析不要用内部数据，券商的分析全部来源外部数据 管理：杜邦分析、敏感度分析 管理：做的如何？用的数据一般都是内部数据——统计报表 (BI) 运营：如何做？ 发展阶段：基础阶段——发展阶段——成熟阶段——优化阶段

	基础阶段 业务数据驱动，纯内部数据	发展阶段 业务数据+数据驱动决策，外部数据+数据驱动	成熟阶段 数据驱动决策，数据驱动决策，外部数据+数据驱动决策	优化阶段 数据驱动决策，数据驱动决策，外部数据+数据驱动决策
数据策略服务	数据以现代方式数字化在流程中	上线数据引擎 业务人员自主配置业务需求和决策流 提供手动解包验证服务	构建数据仓库、策略知识库 策略发布周期1-3天 提供精准灰度发布服务	提供预测自动验证服务 提供策略自学习服务 提供线上自动解包验证服务
数据模型服务	多采用专家模型，少量使用量化统计分析 模型开发、上线周期3-6个月 偶尔的模型校验	大量使用专家规则，少量评分卡技术 提供简单模型的上线服务 模型开发、上线周期1-4月 线下手工模型校验	提供线下自动策略验证服务 提供复杂模型的上线服务 模型开发周期1-4周 模型上线周期1-3天 提供A/B测试服务	提供模型产品、解释异常情况决策替代 支持模型切换 采用自适应算法
数据数据服务	满足统计、决策类数据分析需求 数据标准限于业务部门内部	建立内外融合的基礎数据模型 建立内外融合的基礎数据标准 提供线下数据服务系统 外部数据接入1-3天 外部数据加工周期1-4周 外部数据上线周期1-4周 提供数据基础安全服务	建立模型生命周期管理系统 形成策略、模型数据闭环 提供策略特征自服务工具 外部数据自注册服务 外部数据接入1-3天 建立线上、下一体化数据管控系统的提供数据级安全服务 提供数据进阶安全服务	提供复杂特征自服务工具 实现线上、下数据模型一体化 外部数据自优化服务 建立线上、下一体化数据管控系统的提供数据级安全服务

学习型组织，强调老带新 知识型组织，隐性知识显性化，能说出指标的就是隐性知识显性化、量化、

算法 应聘是要注重聊岗位：企业当前的数据管理处于什么级别，目前处于什么发展阶段。一般互联网公司都是占山为王，各业务部门不愿共享数据，数据缺乏标准化

● 统计报告工具——主要针对管理层决策需求

CDA1重在真实反应历史和当前经营情况

1 常规报表 STANDARD REPORTS
 回答：发生了什么？什么时候发生的？
 示例：月度或季度财务报表
 我们都见过报表，它们一般是定期生成，用来回答在某个特定的领域发生了什么，从某种程度上来说它们是有用的，但无法用于制定长期决策。

2 即席查询 AD HOC REPORTS
 回答：有多少数量？发生了多少次？在哪里？
 示例：一周内各天各种门诊的病人数量报告。
 即席查询的最大好处是，让你不断提出问题并寻找答案。

3 多维分析 OLAP
 回答：问题到底出在哪里？我该如何寻找答案？
 示例：对各种手机类型的用户进行排序，探查他们的呼叫行为。
 通过多维分析(OLAP)的钻取功能，可以让您有初步的发现。钻取功能如同层层剥笋，发现问题所在。

4 警报 ALERTS
 回答：我什么时候该有所反应？现在该做什么？
 示例：当销售额落后于目标时，销售总监将收到警报。
 警报可以让您知道什么时候出了问题，并当问题再次出现时及时告知您。警报可以通过电子邮件、RSS订阅、评分卡或仪表盘上的红色信号灯来展示。

CDA2重在运营经验和算法的交互，模型超参数寻优为交互式

5 统计分析 STATISTICAL ANALYSIS
 回答：为什么会发生这种情况？我错过了什么机会？
 示例：银行可以弄清楚为什么重新申请贷款的客户也增多。
 这时您已可以进行一些复杂的分析，比如频次分析模型或回归分析等等。统计分析是在历史数据中进行统计并总结规律。

6 预报 FORECASTING
 回答：如果持续这种发展趋势，未来会怎样？还需要多少？什么时候需要？
 示例：零售商可以预计特定商品未来一段时间在各个门店的需求量。
 预报可以说是最热门的分析应用之一，各行各业都用得到，特别对于供应商来说，能够准确预报需求，就可以让他们合理安排库存，既不会缺货，也不会积压。

7 流程分析 process optimization
 价值流程图 (Value Stream Mapping--VSM) 是丰田精益制造 (Lean Manufacturing) 生产系统框架下的一种用来描述物流和信息流的形象化工具。价值流程图VSM对生产制造过程中的周期时间、机器运行时间、在制品库存、原材料流动、浪费活动等情形进行测量和记录，有助于形象化当前流程的活动状态，并有利于对生产流程进行指导，朝向理想化方向发展。

8 优化 OPTIMIZATION
 回答：如何把事情做得更好？对于一个复杂问题来说，哪种方案是最优化的？
 示例：在给定工业上的约束条件、资源调配的约束条件以及可用技术的情况下，请您来给出该平台优化的最佳方案，以满足每个用户的需求。
 优化带来创新，它同时考虑到资源与需求，帮助您找到实现目标的最佳方式。

1.常规报表：发生了什么？ 什么时候发生的？ 2.即席查询：发生了什么？ 什么时候发生的？ 在哪里发生的？ 有多少数量？ 80%写 SQL， 蒙牛伊利等传统企业还在使用 Excel， 正在建设数据库。 银行一般用的 oracle， 互联网用的 MySQL 3.多维分析：问题到底出在了哪里？ 该如何寻找答案？ 4.警报 5. 统计分析：回答为什么会发生这种情况？ 哪个方案最高？ 6.预报 7.流程分析 很多策略岗做的就是流程分析和 A/B 测试 8.优化：如何做的更好？ 优化带来创新， 它同时考虑到资源与需求

● 传统企业中的支持性活动逐渐结合数据分析



图：波特价值链

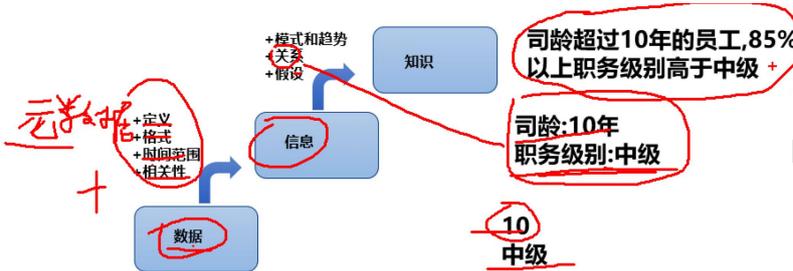


了解企业 ERP 是最快了解业务流程的方式 ERP 就是基于业务流程的固化 数据分析时要了解企业业务流程, 防止用业务流程节点后的数据指导节点前的决策

● 数据治理

为什么要做数据治理? —— “大数据应用” 数据治理问题: 1.如何匹配? 2.匹配效率如何? 3.如何提升匹配率 如保险企业如何和腾讯合作, 推广保险产品 统一客户识别案例: 中国法律漏斗, 某客户有两个证, 一个国内身份证, 一个国外身份证。用国内身份证办办理的贷款逾期了, 用国外的身份证又审批了一个公司又不同的子公司和事业群, 需要统一数据标准, 用统一口径和计算规则核算指标

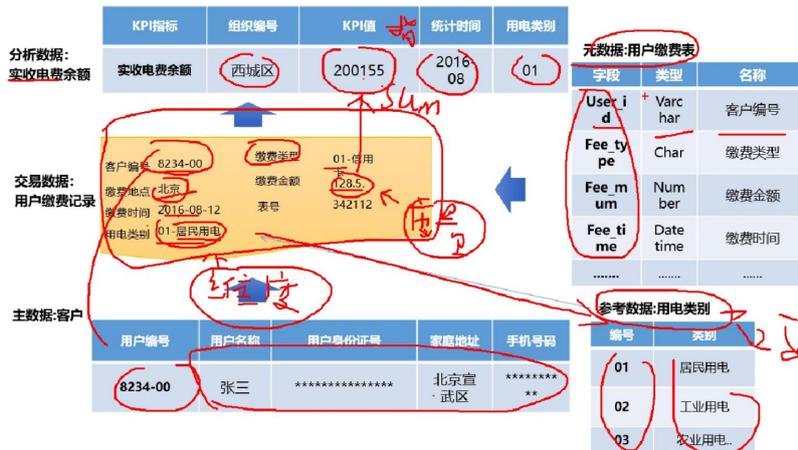
● 数据是对客体的抽象反应



元数据: 元数据就是一张表的表头 元数据+定义、格式、时间范围、相关性>>>信息 信息+模式和趋势、关系、假设>>>信息 知识是规律的近似, 从数据中得到的知识需要不断的检验, 需要做统计分析和假设检验, 不断迭代

● 数据分类

数据分类-数据管理视角示例



1、业务视角 企业就是契约的集合、关系的集合 2、数据管理视角 (1) 主数据 (2) 交易数据 (事务数据) (3) 统计分数 (指标) (4) 参考数据——是维度、分类变量 (5) 元数据——数据字典, 元数据就是数据的户口本 维度和度量 指标就是按照维度对度量进行汇总 前台: 业务 中台: 数据分析, 站在数据视角和业务管理双重视角上 后台: IT 开发

● 数据管理原则: DMBOK2.0

数据管理职能 1、监督: 数据治理 2、数据生命周期管理 计划与设计、使用与增强、启用和维护 3、基础活动 数据质量管理、元数据管理 (让数据看得懂)、数据风险管理 (安全、隐私、遵从)

● 数据治理的组织架构



一、数据治理指导委员会 数据治理理事会: 管理数据标准、质量、安全访问 数据标准管理: 发起——评估——审批——发布——应用——管控——发起 二、IT 组织: 负责落地 三、工作管理 数据管理专员的职责 中国式数据治理组织架构: 牵头业务部门与 IT 部门联合负责, 其中 IT 部门就是执行部门 麦肯锡: 数据治理落地成功率仅 20% 数据标准: 各个业务部门单独是没有动力做的。一家银行有信用卡贷款中心和个人贷款中心, 信用卡贷款中心就想从个人贷款中心中获得客户资源, 但是又不想分享自己的资源, 这时就需要一个强权部门。建设银行总行统计数据仓库, 留数不留人

● 数据治理: 统一标准, 阿里岗位 P6 以上, 属于中高级岗位

● 数据生命周期

数据生命周期（时间角度）



系统开发生命周期 (SDLC)

1、数据生命周期（空间角度） 在线阶段——归档阶段——销毁阶段 数据仓库早期是用来做数据报表的，后来开始做数据挖掘 财务部门：12+1个月存档数据；审计要去：需存储7年 2、数据生命周期（时间角度） 数据规划：需要用哪些数据

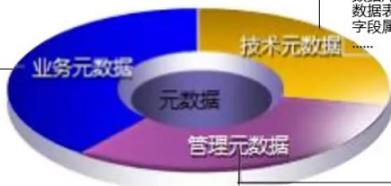
● 企业数据模型

企业应维护一套企业数据模型,作为各项目建设的参考数据模型



面向业务人员,用业务语言从业务术语、指标、规则等几个方面描述数据

业务元数据
业务指标
业务术语
分析应用



面向技术人员,用技术语言从数据库、数据表、字段等几个方面描述数据

技术元数据
数据库属性
数据表属性
字段属性

应用系统
应用编号
归口部门
处理逻辑
.....

库名称
库类型
.....
表名称
表创建时间
主键信息
.....
字段名称
数据类型
.....
数据库记录数
数据库表空间
ETL

面向数据管理人员,从数据管理的视角进行描述

数据治理
数据定义部门
数据生成部门
数据使用部门

数据质量
数据稽核规则
数据质量报告

数据安全
用户权限
数据表授权

数据模型：对数据进行分类并寻找关系 信息价值链分析：使用 CURD 矩阵记录数据在企业不同领域的采集、流动、使用 Create Read Update Delete

数据标准：业务术语标准化、指标标准化、数据模型标准化、数据元素标准化、参考数据标准化 元数据管理：

- (1) 业务元数据：面向业务人员，用业务语言从业务术语、指标、规则等几个方面描述数据
- (2) 管理元数据：数据治理（定义、生成、使用）、数据质量、数据安全
- (3) 技术元数据：面向技术人员，用技术语言从数据库、数据表、字段等几个方面描述数据

- 数据治理管理

数据在用的时候是否满足要求 评估数据质量维度

- 主数据管理：确保多系统的数据一致性，可以共享

- -----企业数据能力建设概述和案例-----

- 数据战略总体设计原则

制定目标——拆解能力（实现目标需要什么能力）——搭建组织结构 数据洞察能力提升：如给客户打标签，需要数字化客户运营数据产品支持（如信用卡中心帮助亲子采摘乐园营销，给什么客户打标签：经常在放学时候出现在学校附近的人）。

- 数据产品

数据产品的表现形式：客户画像、打标签产品

- 数据治理

分清权责 明确治理主体、治理对象、治理手段、支撑系统

- 数据治理的实操框架

数据治理过程：统筹和规划——构建和运行——监控和评价——改进和优化（PDMA）

- 数据治理与数据应用

数据治理的目的就是提高数据质量 数据分析方向：质量分析、风险管控、营销运营（客户管理）、财务管理 数据产品

- 数据管理能力成熟模型（DCMM）

做数据治理就是制度用人，从 PDMA 环节中的 M 开始。先使用量表做诊断 数据分析师是做 M；小组长做 MA

- 数据资产盘点

- 数据安全 CSO

数据分类分级是数据安全管控的基础，让不同的人有不同的访问权限

- 企业架构



当前比较流行的是 TOGAF 一个企业分为前台、中台、后台 前台：用户旅程