

源于社区 服务社区

 中国DevOps社区峰会 2023 · 广州



苏格拉底的申辩 数据治理效果评估困境与应对

神州信息首席数据官 黄万忠



黄万忠

神州信息首席数据官、上地大数据研究院院长
国际数据管理协会DAMA大中华区副主席



曾任Teradata大中华区西区专业服务合伙人，Teradata大中华区数据治理负责人。

获得2019年成都市实体经济新经济领域人才奖励，主导或参与多个监管部门的数据治理研究课题，参与《DMBOK2》、《Navigating the labyrinth穿越数据的迷宫》、《区块链如何改变游戏规则》、《The Unified Star Schema》等书籍翻译。近20年金融行业研发、服务和项目管理经验

专业领域：

数据治理：擅长数据管控、数据资产管理、数据标准咨询、数据安全咨询及其他治理其他领域

数据中台：擅长数据架构、数据模型、数据规划、数据平台咨询和实施

数据挖掘：擅长数据挖掘和机器学习算法，为金融客户提供精准营销、大数据风控、专题分析、隐私计算规划等服务



“我去死，你们去生。
我们所去的哪个更好，谁也不知道，
除非是神。”

alla gar ede opa apienai,
emoi men apothanoumenoi,
umin de biosomenois;
hopotepoi de umon erxontai epi ameinon pragma,
adelon panti plen e toi theoi.



DATA EDGE
数 据 边 缘

Discussion Club
辩论俱乐部

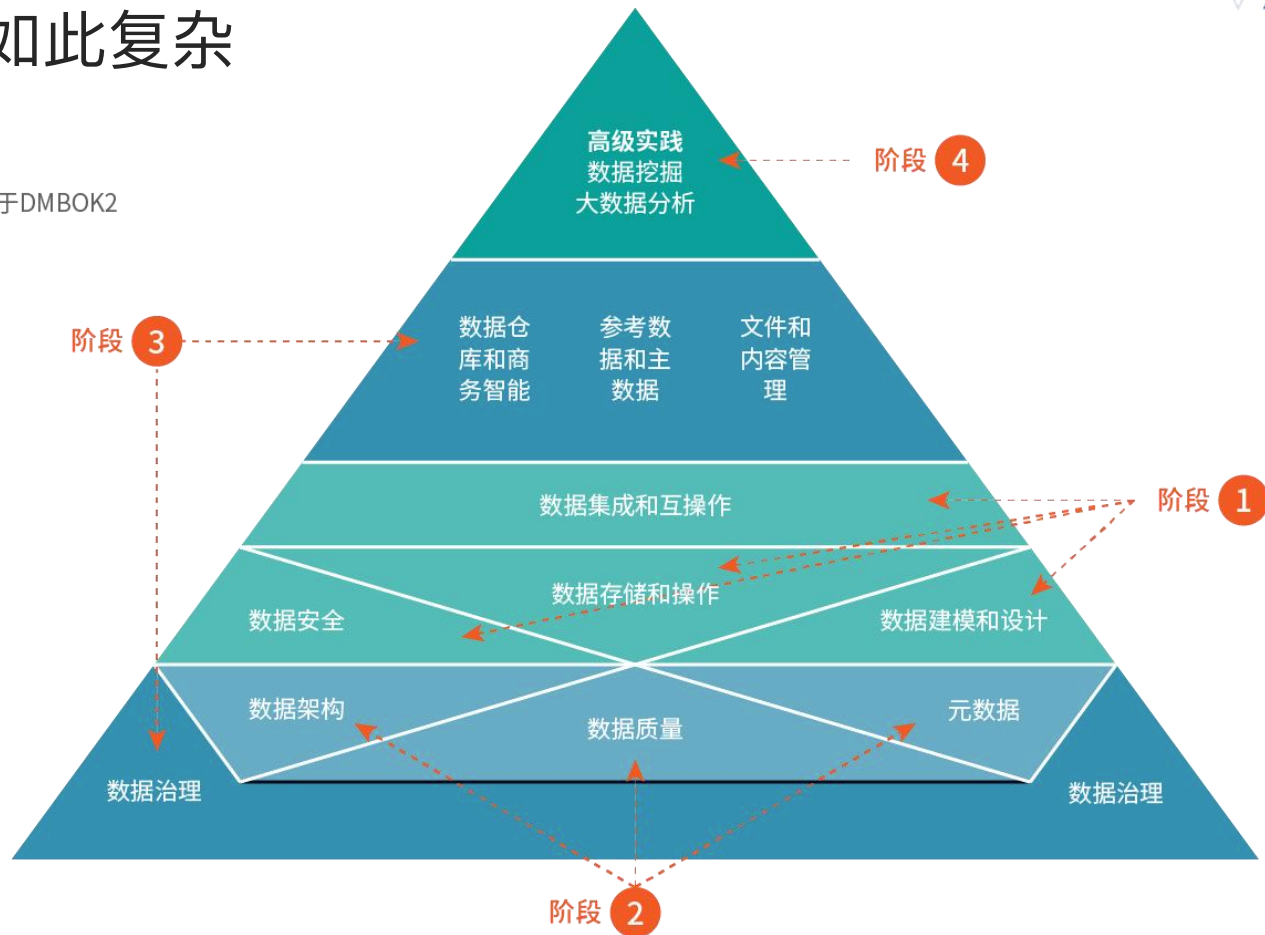
数据治理包括了什么？



- **数据架构管理** 要把水果摆放得紧凑、稳定、便于取放
- **主数据管理** 往往是高价值、受欢迎而放在专区的水果
- **数据标准管理** 水果规格一致
- **数据生命周期管理** 水果保质期管理、下架管理
- **数据安全管理** 防止被人偷吃、损坏
- **元数据管理** 描述水果的品种、价钱和产地
- **数据质量管理** 保证没有烂水果
- **数据服务与价值变现** 导购精准服务

现状：数据治理如此复杂

艾肯金字塔 (Peter Aiken), 来源于DMBOK2



困扰：如何衡量数据治理工作做的是好是坏？

- 做了那么多治理工作，为什么监管报送还要被处罚？

- 领导不太满意，觉得做的不好

- 业务部门没有感觉，觉得对业务没有价值

- 报表出数很慢，得不到想要的指标

- 数据质量还是有那么多的问题

- 数据治理很被动，没有全行级机制

- 只是把数据治理当成一次性活动

- 数据标准落标往往会被妥协

- 总觉得数据治理全都做了，又全都没做好

- 数据治理的热情随着时间在衰退

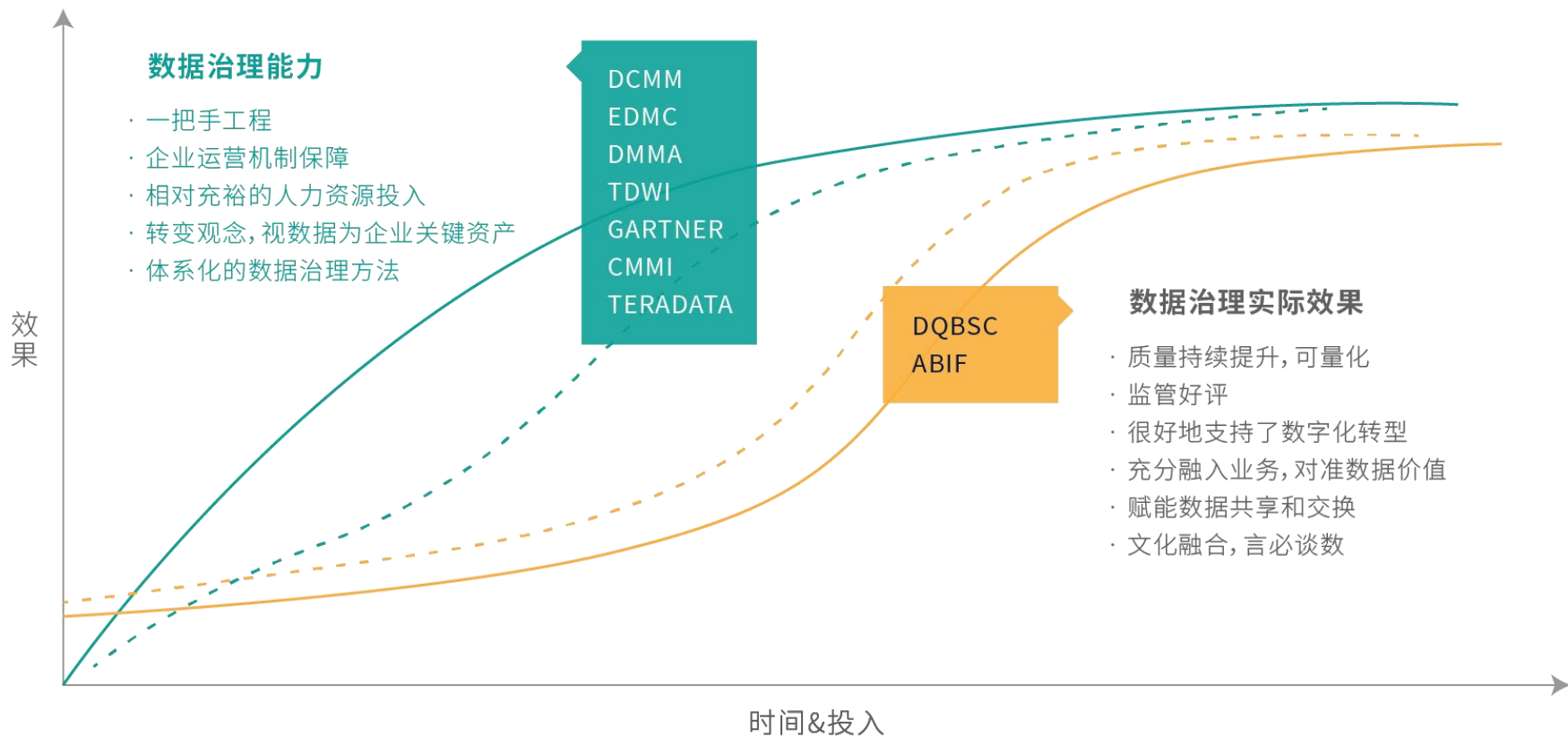
探索：衡量好坏的思路？



1.有没有能力做好？

2.实际有没有提升？

探索：如何衡量数据治理工作做的是好是坏？



探索：数据治理能力成熟度评估

等级	TDW I	Gartner	CMMI	IBM	Teradata	某监管部门	DAMA	DCMM	某监管数据能力建设指引	EDMC
1	孕育	混沌 (Unaware)	初始	初始	启动	起步阶段	初始级	初始级	初始级	未启动
2	婴儿	醒悟 (Aware)	已管理	已管理	理解	被动阶段	重复级	受管理级	受管理级	概念
3	小孩	被动 (Reactive)	已定义	已定义	变革	主动阶段	定义级	稳健级	稳健级	开发
4	少年	主动 (Proactive)	已量化管理	已量化管理	成长	制度化阶段	管理级	量化管理级	量化管理级	定义
5	成年	有效 (Effective)	优化管理	优化管理	竞争	量化阶段	优化级	优化级	优化级	实现
6	圣贤				领先					增强



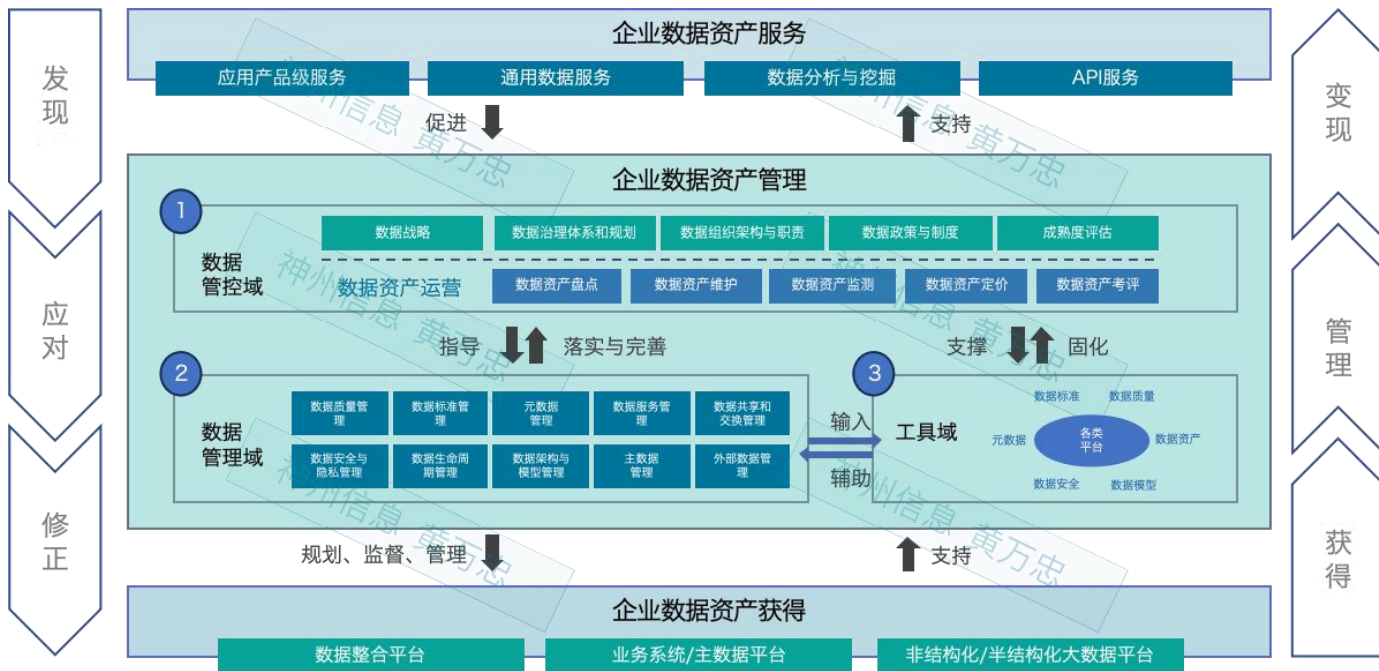
数据治理能力的追求 无知之知的智慧

- 第一类 不知道自己不知道
- 第二类 固执于自己的无知
- 第三类 知道自己不知道
- 第四类 知道自己知道
- 第五类 智者之知
- 第六类 无知之知



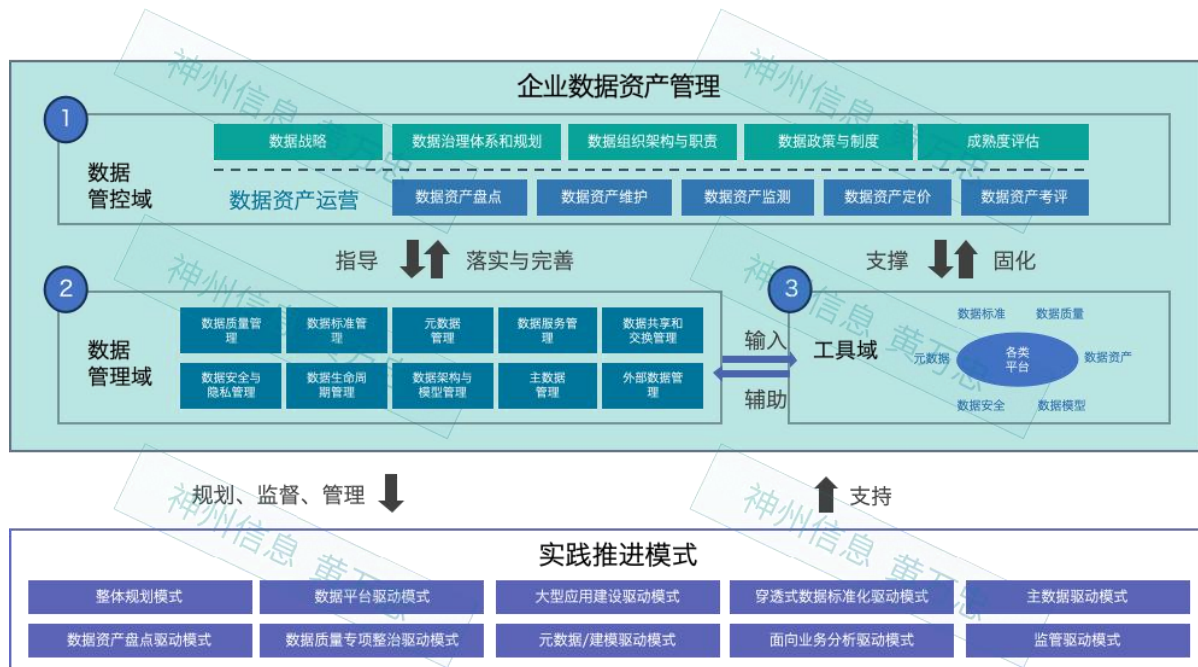
辩一：针对知道自己不知道

专家引导，共同进步，先理论后实践，学习先进案例，由点及面试点，授人以渔，培养数据治理执行团队



辩二：不知无知&固于无知

攻坚一把手，使用敏捷式治理法，快速见效，注意沟通和汇报技巧，避免冲突，保护数据治理执行团队，寻求业务支持



举例：神州信息ABIF数据资产价值提升框架

业务领域	业务提升机会					
客户管理	新增获客 <ul style="list-style-type: none"> -相似客户挖掘 -潜在价值客户挖掘 -公私联动分析 	客户培养 <ul style="list-style-type: none"> -客户行为分析 -客户还款分析 -客户价值提升分析 	客户粘性 <ul style="list-style-type: none"> -用户留存分析 -流失预警分析 -客户体验分析 	客户挽留 <ul style="list-style-type: none"> -客户流失倾向 -客户挽留策略分析 -客户流失原因分析 	客户赢回 <ul style="list-style-type: none"> -传播策略分析 -社交网络分析 -赢回响应策略分析 	生命周期管理 <ul style="list-style-type: none"> -漏斗转化分析 -渠道分析 -客户生命周期
产品管理	市场探索 <ul style="list-style-type: none"> -市场供需分析 -竞品分析 -投资回报分析 	销售推广 <ul style="list-style-type: none"> -产品销量预测 -渠道归因分析 -目标客户分析 	优化升级 <ul style="list-style-type: none"> -产品A/B测试 -用户意见分析 -客户体验分析 	可持续性 <ul style="list-style-type: none"> -用户活跃度分析 -产品价值评估 -客户流失分析 	转型换代 <ul style="list-style-type: none"> -转型方向分析 -转型效果分析 -客户接受度分析 	退出复盘 <ul style="list-style-type: none"> -产品复盘 -竞品对比分析 -市场趋势分析
营销管理	市场分析 <ul style="list-style-type: none"> -消费者需求分析 -市场定位 -SWOT分析 	方案设计 <ul style="list-style-type: none"> -产品营销方案 -差异化定价 -文案内容策略 	渠道投放 <ul style="list-style-type: none"> -渠道活跃度分析 -电子银行转化率 -第三方广告分析 	推广策略 <ul style="list-style-type: none"> -个性化推荐 -关联销售 -推荐算法 	效果评估 <ul style="list-style-type: none"> -响应效果分析 -用户体验分析 -收益分析 	策略优化 <ul style="list-style-type: none"> -产品优化策略 -渠道优化策略 -推广优化策略
运营管理	销售业绩 <ul style="list-style-type: none"> -销售人员绩效 -销售人员聘用模型 -销售奖金激励 	渠道业绩 <ul style="list-style-type: none"> -分支机构-呼叫中心 -终端绩效-网络绩效 -渠道价值模型 	产品业绩 <ul style="list-style-type: none"> -产品绩效模型 -产品定价模型 -产品生命周期模型 	流程业绩 <ul style="list-style-type: none"> -运营绩效管理 -KPI考核与激励 -客户影响分析 	客户价值 <ul style="list-style-type: none"> -客户绩效管理 -客户关键绩效指标 -客户分配 	整合绩效 <ul style="list-style-type: none"> -员工绩效管理 -平衡计分卡指标 -运营状况仪表盘
风险管理	准入策略 <ul style="list-style-type: none"> -黑名单管理 -预授信策略 -监管政策解读 	风险评级 <ul style="list-style-type: none"> -个贷申请评分模型 -小微申请评分模型 -关联方风险模型 	授信策略 <ul style="list-style-type: none"> -审批结论建议 -定价模型 -定额模型 	贷后监测 <ul style="list-style-type: none"> -风险预警模型 -行为评分卡 -穿透式风险管理 	反洗钱反欺诈 <ul style="list-style-type: none"> -申请反欺诈 -交易反欺诈 -反洗钱 	全面风险管理 <ul style="list-style-type: none"> -操作风险 -市场风险
财务管理	监管与披露 <ul style="list-style-type: none"> -银行业监管报表 -财务报表 -财务报告 	资产负债管理 <ul style="list-style-type: none"> -资金转移价格 -资产负债管理 -资产分配引擎 	定价及成本 <ul style="list-style-type: none"> -客户收益分析 -定价模型 -潜在价值分析 	财务预算 <ul style="list-style-type: none"> -财务预算与规划预测模型 	合并与收购 <ul style="list-style-type: none"> -资产负债平衡表分析 -合并重组分析 	SVA绩效管理 <ul style="list-style-type: none"> -股东权益回报分析 -关键指标明细报告
数据管理	数据治理 <ul style="list-style-type: none"> -数据标准 -数据质量 -数据安全 	数据应用 <ul style="list-style-type: none"> -数据建模 -数据集成 -数据共享 	数据服务 <ul style="list-style-type: none"> -数据接口 -数据门户 -数据应用 	数据安全 <ul style="list-style-type: none"> -数据安全 -数据安全 -数据安全 	数据合规 <ul style="list-style-type: none"> -数据安全 -数据安全 -数据安全 	数据价值 <ul style="list-style-type: none"> -数据价值 -数据价值 -数据价值
风险管理	数据治理 <ul style="list-style-type: none"> -数据标准 -数据质量 -数据安全 	数据应用 <ul style="list-style-type: none"> -数据建模 -数据集成 -数据共享 	数据服务 <ul style="list-style-type: none"> -数据接口 -数据门户 -数据应用 	数据安全 <ul style="list-style-type: none"> -数据安全 -数据安全 -数据安全 	数据合规 <ul style="list-style-type: none"> -数据安全 -数据安全 -数据安全 	数据价值 <ul style="list-style-type: none"> -数据价值 -数据价值 -数据价值



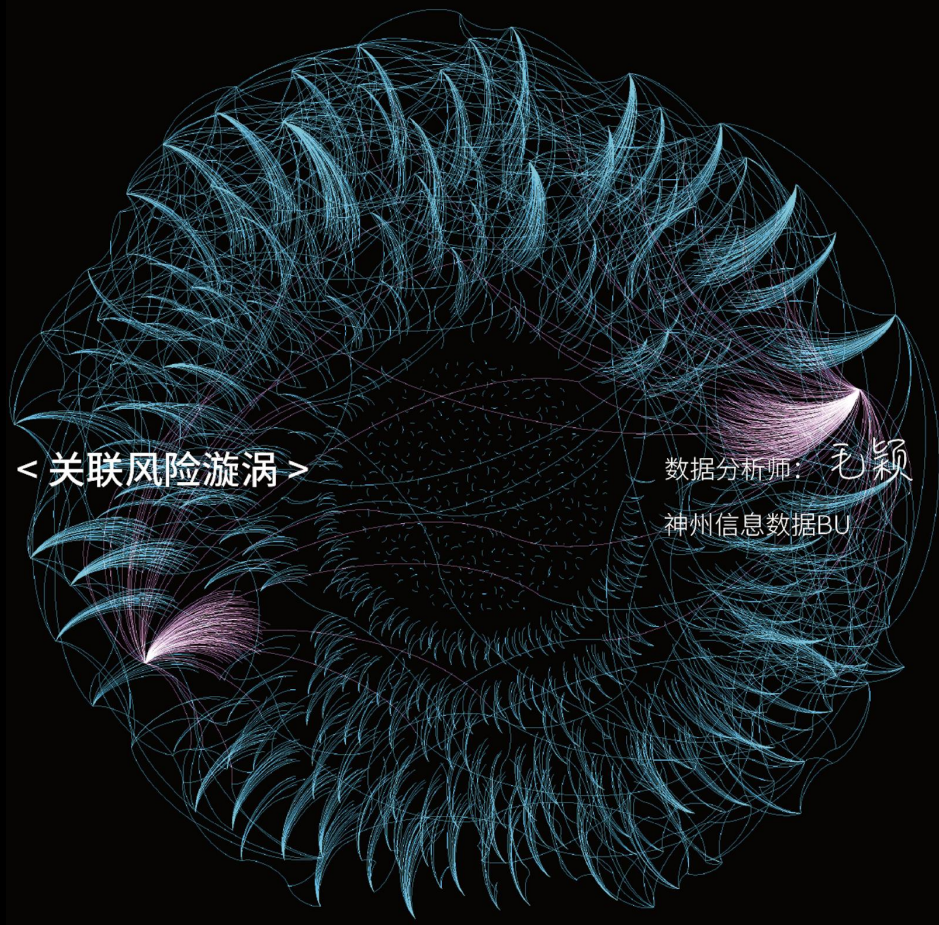
举例：关联风险旋涡

分析场景

银行企业客户间存在大量的交易转账关系，这反映出客户之间的关系紧密程度。当其中一家企业客户发生违约，其代发工资个体极易受到影响。因此需要利用大数据挖掘算法，预测出风险高的企业客户，对其及代发工资客户进行事前严密监控，防范风险。

分析方法

TDA & 基于Stacking的风险预测模型 (XGBOOST & LightGBM & RandomForest)

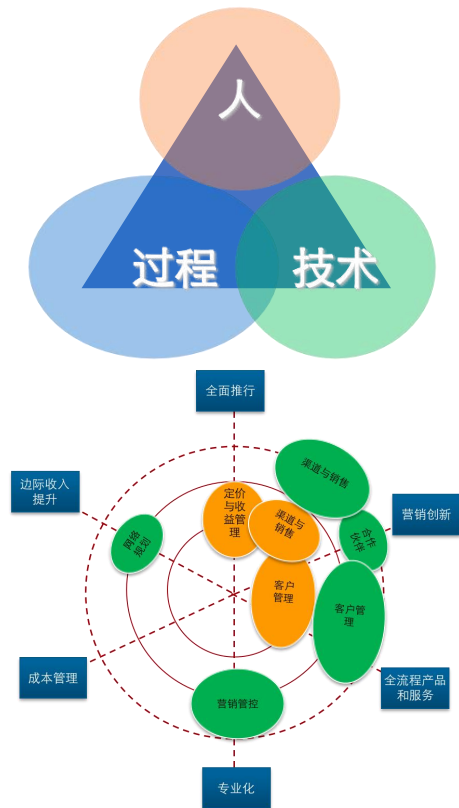


举例：DQBSC



层次	第一层指标					第二层指标					第三层指标				
	加权后综合得分(O) $O = \sum_{k=1}^m O_k \cdot W_k$	指标名称	指标权重(W _k)	指标满分值(T _k)	实际得分值(K _k)	完成率CR _k (%)	指标名称	指标权重(W _t)	目标	指标原始得分(T _t)	指标原始得分修正系数(C _{ct})	加权后指标得分(K _t)	完成率CR _t (%)	计算公式	
1	$K = \sum_{k=1}^m K_{sk} \cdot W_{sk}$ 备注：m表示参与计算的下层二级指标个数	映射率指标 (Mapping Rate)	20% (可以依据权重设置调整)	$T_s = \sum_{i=1}^n T_{ts}$ 备注：n表示下层二级指标的个数	$K_s = \sum_{i=1}^n K_{ts}$ 备注：n表示下层二级指标的个数	K _s /T _s	CK_MAPP_001(原始映射率)	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t	$K_t' = (\text{已映射数} / \text{需映射总数}) * 100$	$K_k = K_t' * C_{ct} * W_k$
2							CK_MAPP_002(字段映射率)	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
3							CK_MAPP_003(记录映射率)	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
4							CK_MAPP_004(字段映射率)	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
5							CK_MAPP_005(代码映射率)	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
6							CK_MAPP_006(代码映射率)	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
7		完整性指标 (Completeness)	20% (可以依据权重设置调整)	同上	同上	同上	CK_COMP_001	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t	$K_t' = (\text{完整记录数} / \text{记录总数}) * 100$	$K_k = K_t' * C_{ct} * W_k$
8							CK_COMP_002	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
9							CK_COMP_003	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
10		唯一性指标 (Uniqueness)	15% (可以依据权重设置调整)	同上	同上	同上	CK_UNIQ_001	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t	$K_t' = (\text{不重复记录数} / \text{记录总数}) * 100$	
11							CK_UNIQ_002	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
12							CK_UNIQ_003	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
13		有效性指标 (Validity)	15% (可以依据权重设置调整)	同上	同上	同上	CK_VALID_001	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t	$K_t' = (\text{有效记录数} / \text{记录总数}) * 100$	$K_k = K_t' * C_{ct} * W_k$
14							CK_VALID_002	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
15							CK_VALID_003	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
16		一致性指标 (Consistency)	20% (可以依据权重设置调整)	同上	同上	同上	CK_CONS_001	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t	$K_t' = (\text{一致记录数} / \text{记录总数}) * 100$	$K_k = K_t' * C_{ct} * W_k$
17							CK_CONS_002	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
18							CK_CONS_003	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
19		及时性指标 (Timeliness)	10% (可以依据权重设置调整)	同上	同上	同上	CK_TIME_001	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t	$K_t' = \text{Time_Estim}(T_{down} - T_{up}) / T_{up}$ ，该公式说明见备注1	$K_k = K_t' * C_{ct} * W_k$
20							CK_TIME_002	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
21							CK_TIME_003	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
22		准确性指标 (Accuracy)	0%				CK_ACCU_001	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t	暂时无法自动计算	
23							CK_ACCU_002	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
24							CK_ACCU_003	根据权重参数计算	100 * W _t	系统计算	建议设置为1	系统计算	K _t /T _t		
25															
52		数据完整性	0%				CK_VCD_001							数据完整性=1	
53		数据准确性	0%				CK_VCD_002							数据准确性=1	
54		数据一致性	0%				CK_VCD_003							数据一致性=1	
55		数据及时性	0%				CK_VCD_004							数据及时性=1	

神州信息数据治理效果评估苏格拉底框架V1.0



			0	1	2	3	4	5	6	7	8	9	10
人	管理者	五级“无知之知” 管控法											
	执行团队				●								
	直接受益者					●							
过程	启动	PEE评估法					●						
	执行	DEE评估法			●								
	收尾	CEE评估法				●							
	运营	AEE评估法					●						
技术	治理能力评估 (DCMM、EDMC等)							●					
	期望收敛法EC								●				
	DQBS						●						
	ABIF					●							
价值	直接价值评估法									●			
	间接价值评估法									●			





唯有通过认识自己的局限 来思考、演进、优化、辩证数据治理的方法



源于社区 服务社区

THANKS!

