

数据以简为美



指标中台建设之路



痛点在哪?



集团数字化战略

每个BU都在搞自己的 "数字化经营平台" 数据需求暴增

数据治理之痛



懂业务、懂数据、懂技术

上能写制度办法,下能查数据分析



懂权术、懂策略、学兵法

因为具体的工作八成都是跟人有关



最后沦落成wiki,无法持之以恒

治理到一半,前面的部分又开始乱了 制度过于复杂,没有团队能坚持



- 反向型治理,不一致;
- 被动型治理,不高效;
- 误区型治理,难延续;
- 兼职型治理;难落地。

数据治理是场马拉松 , 也是场接力赛。

我部的优势





经过现状分析,结合业界&历史教训

我们提出了场景化治理概念,提倡边应用边治理从指标治理先切入

银行BI体系演化

- 1、报表成体系化建设
- 2、数据即对象,对象即数据
- 3、培养用户统一看数据习惯

智能诊断、方法推荐、跟踪评估 形成大经营闭环

智慧经营时代



算法 数据沉淀 业务场景

数据头条时代

指标、机构

行员 产品











堆砌报表时代

- 1、报表堆积如山
- 2、指标口径打架
- 费时费事费人
- 重复劳动严重
- 5、团队没有成长



表哥 表姐

指标中台建设之路



顶层简单

简单的接入形式

- 数据开发不用关心维度,只需按规范接入,在宽表上插一列。维度关联,衍生数据全自动搞定。
- 从埋点系统、活动平台、标签等系统批量同步指标和维度

强大的技术支撑

从数据模型到平台架构,对现有业务进行 充分的调研评估。踩坑后迅速做出应变, 走一步看百步。

完美的产品设计

全程贯彻所见即所得的设计理念,让开发和业务都愿意接受这个产品,工作模式的革新。

丰富的应用场景

- BI化 --- 取代并收口所有报表看板 类需求,让业务自给自足。
- AI化 --- 让分析类需求场景化,降低分析门槛,人人都是分析师。

持续的数据治理

从指标及维度两方面对口径、质量、时效构建治理方法论。用大量的自动化方法去发现问题,快速治理。

强大的执行力

设计、研发、沟通、推进、检查、推广、 运营、决策、迭代。产品、研发、数据、 测试、PMO团队齐心协力。



潘多拉指标平台 —— 提供—站式数据服务

- 1分钟找到数据
- 3分钟创建看板
- 5分钟挖掘价值

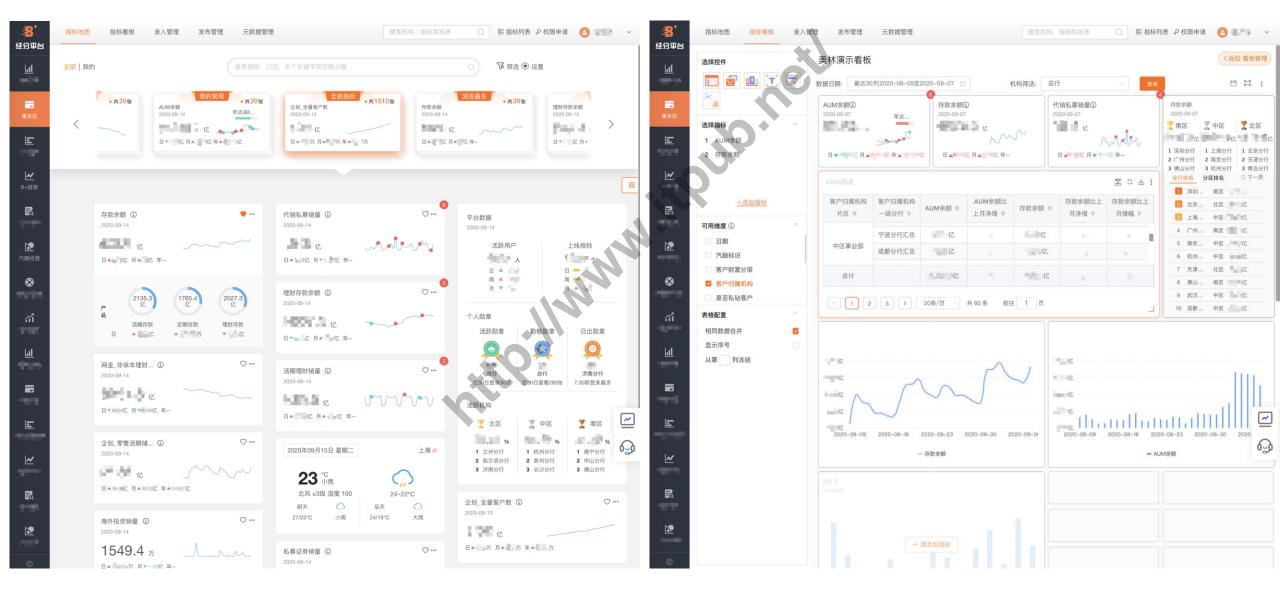


客群分析 成本管控

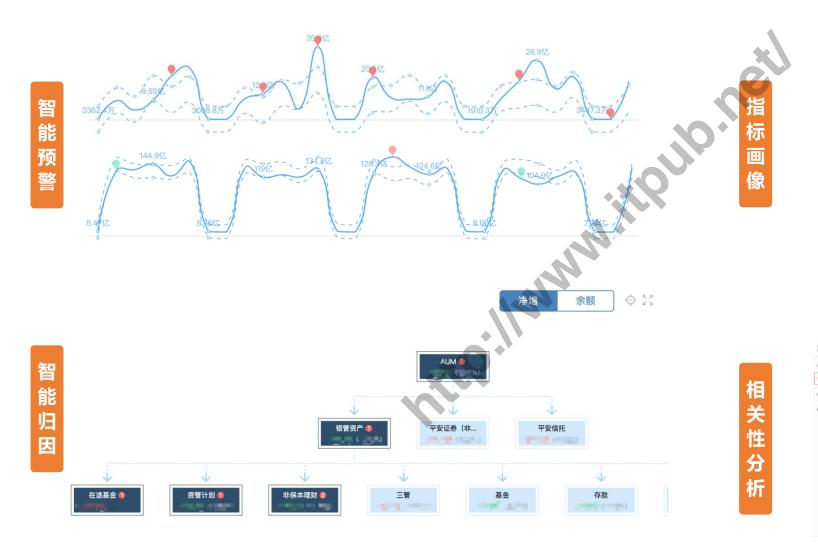
✓ 营造数据生态

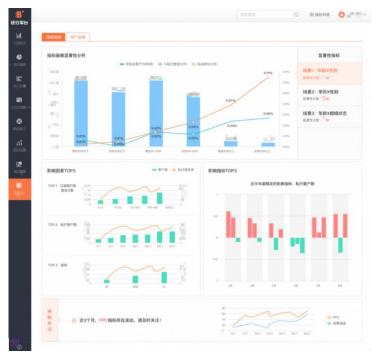
- √ 数据资产治理
 - ✓ 降成本提产能

指标BI化



指标AI化







指标中台建设之路



更重要的是,我们需要专业的技术团队支持,所以我们选Kylingence

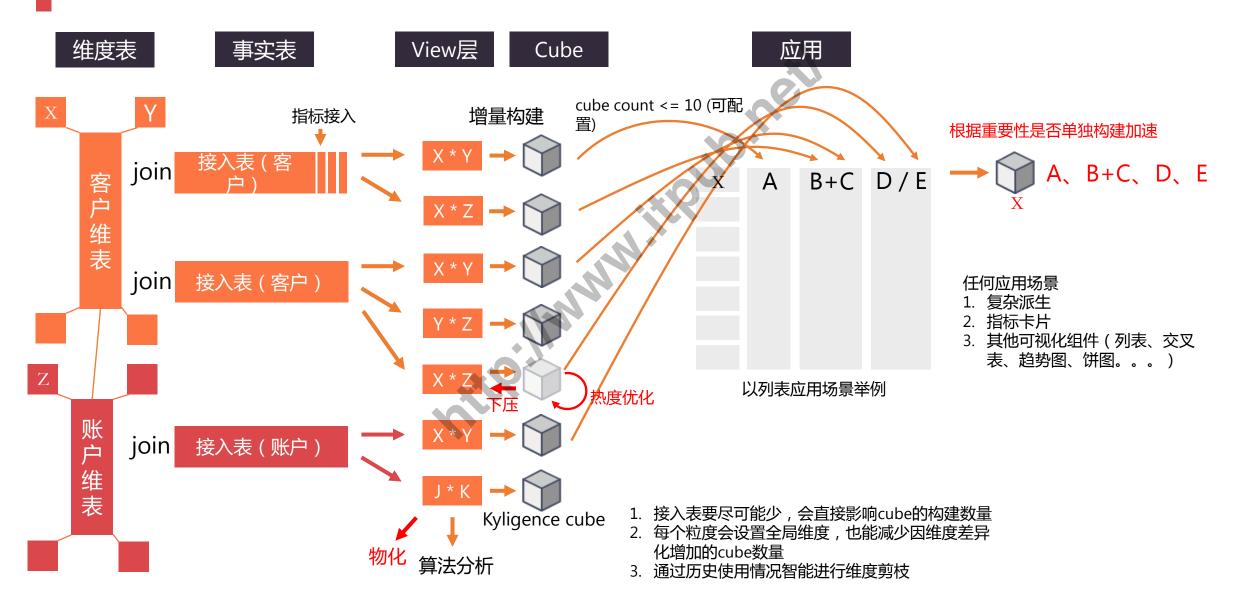
为什么选Kyligence

在对原生hadoop的支持、对高QPS场景的支持、无缝集成主流的 BI 平台等方面, kyligence的优势要高于其它竞品 (Druid ,Clickhouse);

	高QPS	巨大数据量	秒级查询响应	灵活查询方式	Hadoop 生态兼容性
Hive	×	✓	×	/	√
Spark	×	√	×	✓	√
Impala	×	√		✓	✓ (必须部署CDH环境)
Druid	× (非时间维度下预计算)	√	✓ (非高基维)	×(不支持标准SQL)	×
Clickhouse	×(非聚合查询)	×14	✓ (非高基维)	√	×
Flink	×	→	×	✓	√
Kyligence	√	***	✓	✓	✓

- 1. Hive、Spark、Impala、Clickhouse、Flink都是实时聚合查询,无法支持高QPS
- 2. Druid必须在时间维度下做预计算,才能实现高QPS和秒级响应,对非时间维度下、高基维度下的聚合查询效果不佳,且不支持标准SQL语法,实现成本高
- 3. ClickHouse更擅长明细级别的高性能查询,在复杂或者高基维的聚合查询中,无法同时满足高性能和高QPS
- 4. Druid和ClickHouse无法兼容原生Hadoop,需要额外搭建集群存储和处理数据

模型架构



平台架构

指标应用 相关性分析 指标看板 智能预警 归因分析 API / MDX 查询服务 权限校验 微服务 异步查询 度量合并 队列管理 行列转换 数据预览 降级管理 逻辑层封装 查询总线 主/被动缓存

构建服务 视图构建 变更触发 Cube构建 元数据回写 笛卡尔积 数据校验 判断 任务管理 重跑管理 资源检测 接入适配

元数据 事实表管理 维表管理 派生/衍生 血缘分析 影响性分析 Scheduler

监控服务

断点检测

构建监控

日志分析

SQL封装

质量监控

资源监控

自动发现

埋点采集

一致性检测

配置库

指标搜索

轻度聚合

指标计算

Neo4j

Redis

Kyligence (核心)

Spark/Presto

调 度 平

元

数

据

平

台

产品架构

指标应用



指标管理



需求接入



经营分析生态架构一为什么说是中台

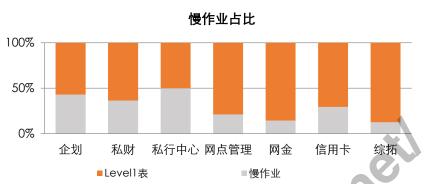
依赖系统 联动系统 场景 营销分析 经分看板 业绩检视 客群分析 成本管控 客户旅程 调度平台 用户行为分析 标签平台 组件 监控预警 -智能推送 其他 指标分析 可视化 客群分析 算法平台 报表配置 智能预警 相关性分析 明细下载 画像分析 标签平台 报警策略 指标卡片 客群对比 指标预测 数据填报 资产平台 数据补录 效果评估 自助报警 自助分析 归因下钻 机器学习平台 实时数仓 指标平台 离线指标 实时指标 即席查询平台 欧拉图谱 数据 资产平台 API 业绩数据 用户行为 活动数据 成本数据 KPI数据 监控平台 口袋银行家 行员标签 人事数据 客户标签 产品标签

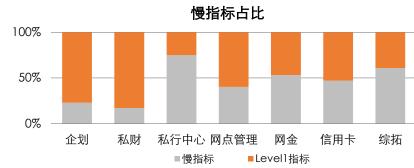
指标中台建设之路



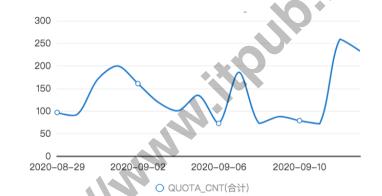
治理方法论

时效治理 口径治理 质量治理 重复指标 指标 断点指标 **SLA Level** 异动指标 Level1: <=8:30点 名称规范 Level2: 8:30~10:30点 Level3: 10:30~15:30 点 口径规范 Level4: 15:30点后 重复维度 维 度 空值属性 分布异常





异 常 指 标 检 测



数据日期	业务线	异常类型	指标名称	近30日异常》	指标汇总值	开发人	前一日	前日汇总值
2020-09-13	网点管理中心	无数值	运营_预审时长	30	<null></null>	EX-WANGGUA	20200912	<null></null>
2020-09-13	网点管理中心	总值为0	经过首页面业务	28	0.0000	EX-WANGGUA	20200912	6193.0000
2020-09-13	网点管理中心	无数值	运营_预约时长	21	<null></null>	EX-WANGGUA	20200912	<null></null>
2020-09-13	网点管理中心	大幅波动	运营_设备服务师	13	2.0000	EX-ZHANGLIJ	20200912	7.0000
2020-09-13	网点管理中心	大幅波动	运营_厅堂设备?	12	17.0000	EX-ZHANGLIJ	20200912	51.0000
2020-09-13	网点管理中心	大幅波动	运营_厅堂设备?	12	0.0024	WEIYANFANG	20200912	0.0140
2020-09-13	网点管理中心	大幅波动	运营_厅堂设备?	12	8.0000	EX-ZHANGLIJ	20200912	108.5000
0000 00 40	TO HARTEL A	24 P±3L ^	\-# \n & nn & c	**	0.0000	EV 711110111	00000040	0 0000

里复指标检

210 180 150			
90 0	/		
60			
30			
0 —			
2020-08-29	2020-09-02	2020-09-06	2020-09-10
		〇 重复指标组数	

20000	理が自動	THEOLIN	三多月日か	H101010101	TESS2010/2522200
信用卡中心	GAT SHOP THE	78785972.0000	78785972.0000	ALC: No.	信用卡中心
零售经营及消费贷款事业部	end Albertania	664358479997.0976	664358479997.0976	this probability.	零售经营及消费贷款事业部
零售经营及消费贷款事业部	distant.	681025065527.0484	681025065527.0484	In James 1919	零售企划
信用卡中心	Section and the	8119368.3100	8119368.3100	Min. printers.	信用卡中心
官用卡中心	SERVICE STREET	88165.4500	88165.4500	975 CO.	信用卡中心
网点管理中心	Character Service	6976052.0000	6976052.0000	May British Salarages	网点管理中心
网金及财富管理事业部	ett (district)	204147378226.5000	204147378226.5000	STATE OF THE STATE	网金及财富管理事业部
从钻客群事业部	High.	9295969884.4700	9295969884.4700	profittigittiget	网金及财富管理事业部
网金及财富管理事业部	years and the second	149120502827.2900	149120502827.2900	Simulating.	零售企划
言用卡中心	gradients	1037335.7200	1037335.7200	who, expense	信用卡中心
网金及财富管理事业部	Contribution and a	1835715126415.4200	1835715126415.4200	State Commence State Co.	~ // 富管理事业部
零售经营及消费贷款事业部	(Aug Malays	664358479997.0976	664358479997.0976	A - PER SE	零售经营及消费贷款事业部
官用卡中心	State (p)	9156704.0300	9156704.0300	will play to	信用卡中心
官用卡中心	Professor Aller	23264174.4700	23264174.4700	400	信用卡中心
官用卡中心	Service Service	1638036.6100	1638036.61	ert-profesite	信用卡中心
官用卡中心	(Contract)	88165.4500	88165.4500	error of the	信用卡中心
信用卡中心	94100acc59	31748036.0000	31748036.0000	ALC: Or course	信用卡中心
零售企划	distribution and the second	681025065527.0484	681025065527.0484	March 1987	零售经营及消费贷款事业部

指标值	重复指标值
78785972.0000	78785972.0000
664358479997.0976	664358479997.0976
681025065527.0484	681025065527.0484
8119368.3100	8119368.3100
88165.4500	88165.4500
6976052.0000	6976052.0000
204147378226.5000	204147378226.5000
9295969884.4700	9295969884.4700

维度治理

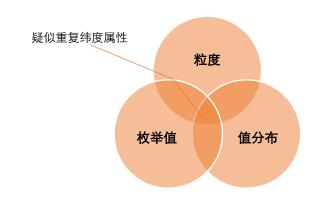
重复纬度

空值属性

纬度分布异常

值治理

接入纬度时进行 校验,发现异常



字符型或日期	理字段统计分析		2	i di	11/1
序号	字段名称 🌲	中文名称 🌲	字段类型 💠	整战 ♦	(i
		10	*		
22	GH_CHANNEL	管户渠道	STRING	100.00%	2
23	GH_POSITION	客户管理人对应的职位	STRING	0.00%	1



纬度名称

名称治理

制定规范,定期 整改

- 1. 接入的纬度实体给出一个实体名称
- 2. 纬度展示名称 = <纬度实体名称>的<属性名称>

举例,接入的客户纬度表中包含如下属性和标签。

字段描述。	属性/标签。	展示纬度名称。
客户财富分层。	属性。	客户的客户财富分层。
是否集团员工。	标签。	客户是否集团员工。

纬度路径别名

渠道粤(6) 优化前

客户号 > 客户管理人 > 所属团队 > 渠道号

优化后 → (别名) 客户管户人所在渠道

规则:

- 给纬度外键取别名
- 发布时纬度展示: <外键别名+的>*+字段原生名字
- 路径节点<=3,直接展示

举例:客户号>客户管理人>所属团队

- → 客户的管理人的所属团队
- 路径节点>3, 需用户给路径给出别名

举例:客户>客户的管户人>行员的所属团队>团队的渠道

→ (别名) 客户管户人所在渠道

指标中台建设之路





边应用边治理

内容是应用的关键

资源优化

减员增效

数据治理

场景复制

上线应用情况

项目研发从19年12月到7月正式上线(中间经历过年和疫情) 上线3个月内接入情况如下:



显著的效率提升

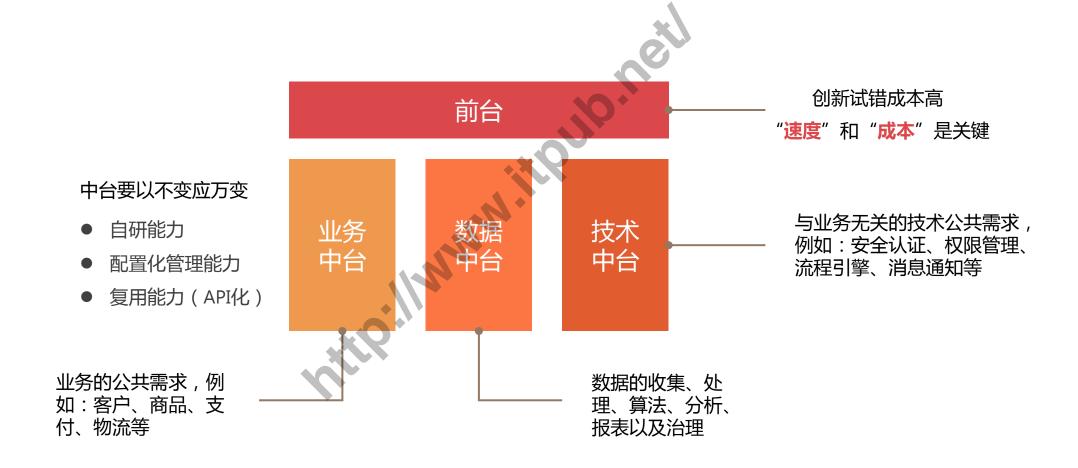
存在数据孤岛,团队 之间指标不通不共享 重复开发严重 业务 业务 其他线IT 传统报表需求 确定要开 提出50个指标 发30个 报表制作 数据探源 数据开发 上线验收 需求提出 需求澄清 1st day 2st day 3st day 10st day 12st day 6st day 开发30个指标,25个重复开发 耗时12个工作日 效果对比 牵扯>3方协作 数据未实现前自动模拟数据 报表直接可预览 指标中台支撑后 只需5个指标,仅耗时5个工作日 创建50个指标 其中30个已有 只需2方协作 并沉淀20个指标 同时拖拽报表 其中10个派生 需求提出 (通过系统) 其中10个需开发 其中5个高级派生 1st day 上线验收 沉淀20个指标 需求澄清 5st day 2st day 其中5个需研发

沉淀5个指标

3st day

师能复用这次沉淀果实

数据中台的价值







大数据平台研发团队出品

