



释放数据潜能 驱动业务增长

WeData数据治理介绍及内部实践

腾讯云大数据平台产品中心 fairyxwang 202208





- 1. 数据治理挑战
- 2. 腾讯内部数据治理实践
- 3. WeData数据治理平台能力

# 01

## 数据治理挑战

## 数据治理挑战













#### 数据信息分散

- 业务系统不清晰
- 数据资产不明晰



#### 元数据不全

- 需要有业务接口人维护 补全业务信息
- 数据多样化,缺少统一 标准



#### 数据质量差

- 上游是一张空表,或者 无效字段,总靠人肉来 判别
- 修改一张表,不清楚带 来哪些影响



#### 维护困难

- 数据有问题,总是事后 才知道
- 不知该由谁去维护预防



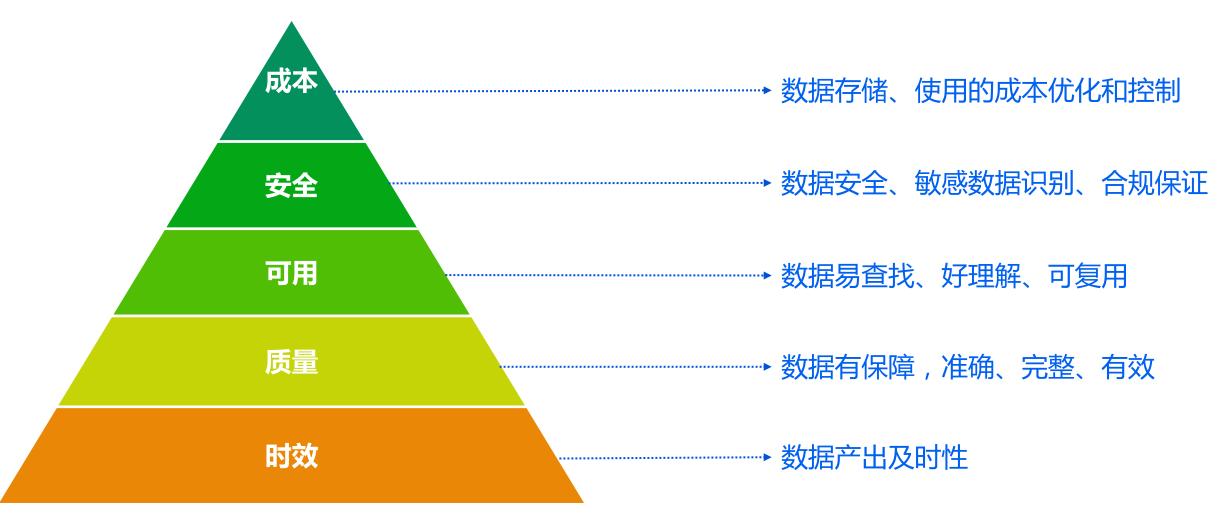
#### 无法审计和度量

- 用数据的人这么多,如 何保障公司数据访问是 可靠安全的
- 数据成本是什么样





### 数字化不同阶段,数据治理关注的核心需求不同



## 02

## 腾讯内部数据治理实践

## 腾讯内部业务现状









## 腾讯数据治理三阶段







## 数据资产化

全面数据汇聚

统一数据资产

业务决策辅助

基于腾讯统一的数据平台体系, 打通系统间 数据孤岛,建立统一的数据资产,为辅助业 务决策提供高质量的数据保障



### 降本增效

成本洞察

成分分析

责任到人

业务规模日益增长,同时也带来巨大的成本 负担,成本治理迫在眉睫



方法论沉淀

治理工具化

效果可量化

将治理方法论通过平台进行实现,治理流程 工具化,提升治理效率,推动深度治理

## 腾讯内部实践:腾讯新闻数据资产化





#### 背景与目标

- 缺少统一数据规范:各业务数据埋点规范、上报规范、 数仓规范、指标规范各异。
- 数据质量难以保障:业务数据仓库庞大,总离线表数超过 xx万,缺乏数据分层及数据模型,数据复用度仅15%, 存在大量年久失修的数据

过程:数据模型重构

埋点设计 ->数仓建模 ->指标建模

统一埋点模型

升级数仓模型

构建指标模型

#### 成果

腾讯新闻数据治理项目目标是通过对腾讯新闻业务数据的 规划、生产、加工、应用等进行全链路改造与升级,在提 升新闻数据的准确性、覆盖度和实时性的同时,打造科学 的数据应用体系。将数据质量体系在新闻整个业务推进落 地,为敏捷分析与科学实验决策等奠定了基础



完成数据生产链路的规范化建设,从埋点到数仓到指 标等,梳理完成了适合新闻的管理流程,并在大改版 过程中快速应用



完成250个模型设计或重构,52个维表的设计以及270 个应用表的开发。在数据资产完整性和分层规范 达到 95% 复用度达到73%以上,跨层引用占比小于 5%

## 腾讯内部实践:PCG数据成本治理





#### PCG大数据成本



#### 资源用量



#### 资源单价

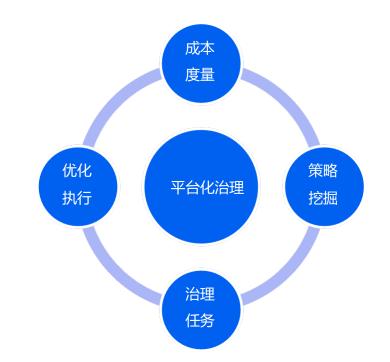
#### 范围:

- 1. 数据采集平台
- 2. 数据生成平台
- 3. 数据分析平台
- 4. 数据应用平台

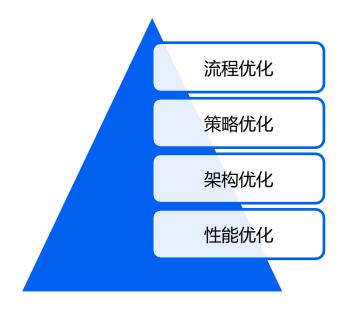
#### 成果:

- 在月成本同比增加30%+的情况下,业务单位用户/内 容消费的大数据成本下降,业务大数据成本绝对值下降 至少10%
- 强化大数据成本治理理论,沉淀方法论、流程和平台能 力

优化思路:降低业务不合理使用



优化思路:提升数据平台效能



## 腾讯内部实践:治理平台化推动业务治理落地





#### 背景与目标

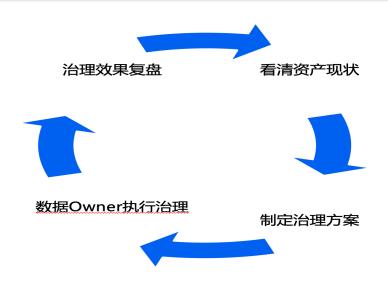
大数据环境下,海量数据给业务带来了巨大价值,同时 也带来巨大的成本及负担。业务团队大数据成本盘点困 难、治理执行门槛成本双高、治理效果不能有效量化, 都是业务在推进资产治理的痛点

#### 过程:构建一站式数据治理平台

现状概览	资产明细	治理方案	治理执行		
洞察全域数仓及大数据成本,呈现最真实的业务资产情况	全方位治理项洞察, 基于资产分规范, 对当前业务的待治 理项进行明细展示	平台内置治理方案 结合业务定制化治 理方案	一站式治理执行, 针对洞察出的治理 项进行治理动作		

#### 资产价值评分体系构建:规范、安全、质量、成本、应用

资产指标维度	指标名称	指标引导目的	指标计算口径	不及格分标准(0分)	60分标准	80分标准	90分标准	100分标准	单项机
2 2	1)表名命名规范性指标	表命名更加规范	a)表模型层级前膜不符合规范 b)表后缀命名不符合规范 c)表主题命名不符合规范	命中a,b,c	命中b,c	命中c或者b	命中c	无任何命中	x9
	2)表和字段注释规范性指	表注释更加规范易懂	a)表没有注释 b)表有注释,但是注释太短 c)字段没有注释 d)字段部分有注释,或者有注释的字段占比(目前定 90%)	命中a	未命中a,但命中b 或c	未命中a,b,c, 只命中d	-	无任何命中	×9
	3)字段安全等级性指标	表权限管理符合规范	a)该表字段有安全等级配置占比<10% b)该表字段有安全等级配置占比<40%,但>=10%; c)该表字段有安全等级配置占比<60%,但>=40%; d)该表字段有安全等级配置占比<80%,但>=60%;	命中a	命中b	爺中の	命中d	无任何命中	x9
	4)责任人符合规范	表责任人合理有人管	a)表扫属离职人员 b)表bu和owner的BU不一致 c)表owner和任务生产owner不一致 c)表owner和最近一次DDL的owner不一致 信务元数据不准确。暂度异)	命中a	命中b,无a	命中c, 无a,b	命中d, 无a,b,c	无任何命中	×9
	5)依赖加工规范	引导表依赖和计算加工更加规范	ai定居依赖,连背了数据模型主体流程依赖的顺序。 0%-2dwd-2ww-2dslwin的顺序,出现直接或者跨 多级逆序的; b)底层依赖,出现除dwd以外依赖也局的,如dws 依赖dots,adslwin依赖des的; c)加工链路过长,数仓模型中,从源头别该表的任务 层架补的(是机分编取例是不和分);	命中a	未命中a,命中b;	未命中a, b,命中c; (调整为不扣分,即此 条规则作废)	-	无任何命中	×9
保障性评估	1)有监控保障指标	引导任务有监控管理	a)表没有任何监控 b)教仪有非强制性监控,且监控数=1 c)表仅有非强制性监控,但监控数>1 c)表有强制性报错监控,且监控数=1 e)表有强制性报错监控,且监控数>1	命中a	命中b	命中c	命中d	命中e	×
	2)有DQC报警保障指标	引导资产产出有DQC保障	a)表没有报警owner配置 b)表没有现信、企復报警配置 c)表没有电话报警配置 o)表没有电话报警配置 o)表没有产出承诺配置	命中a	没有命中a,命中b	没有命中a,b, 命中c	没有命中a,b,c,命中d	无任何命中	×
	3)有基线保障指标	资产生产有资源或承诺保障	a)表没有任何基线 b)表有普通任务基线 c)表有高优先级任务基线	命中a	-	命中b	-	命中c	×,
准确性评估	1)DQC达标指标	引导数据产出准确	a)最近一天出现数据为空报警 b)最近一天出现数据重复报警 c)最近一天表符产出现强阻塞报警	£-та.о.о.ц. го				ALTERT	



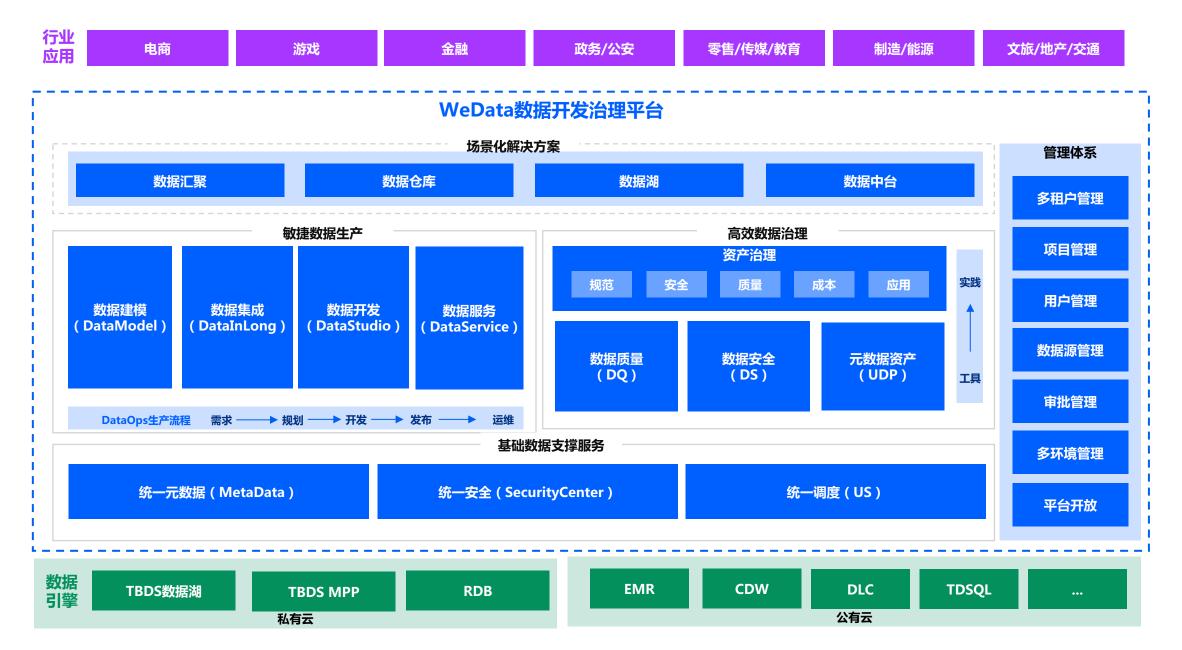
# 03

## WeData 数据治理平台能力

## 腾讯内部大数据能力的对外商业化输出-WeData





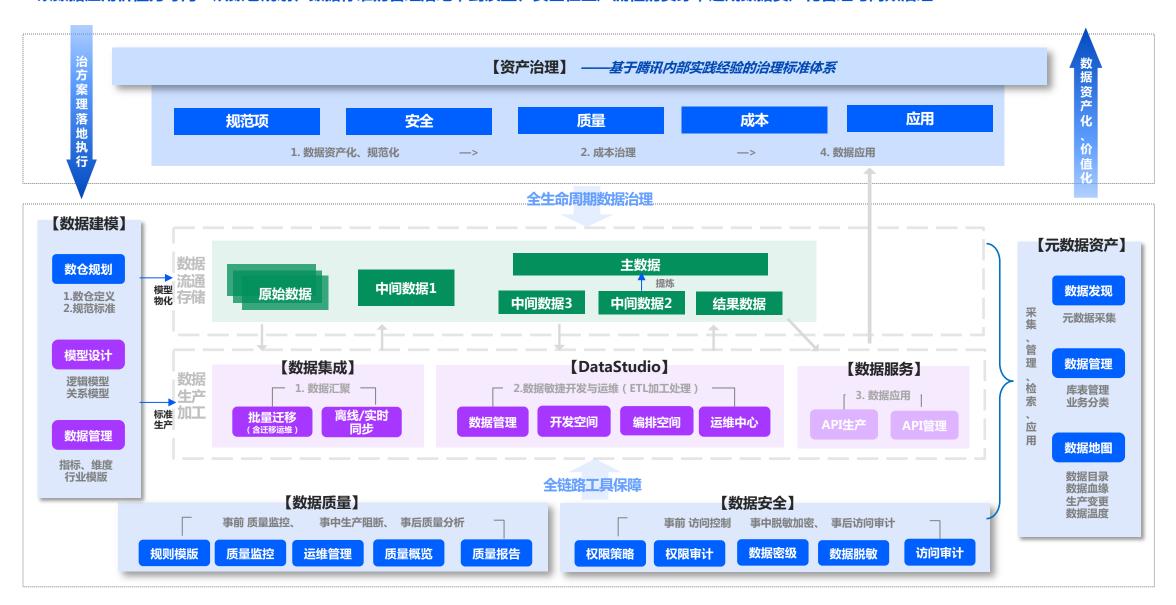


## WeData数据治理服务





以数据应用价值为导向:从数仓规划、数据标准的管理落地,到质量、安全在生产流程的贯穿,达成数据资产化管理与高效治理

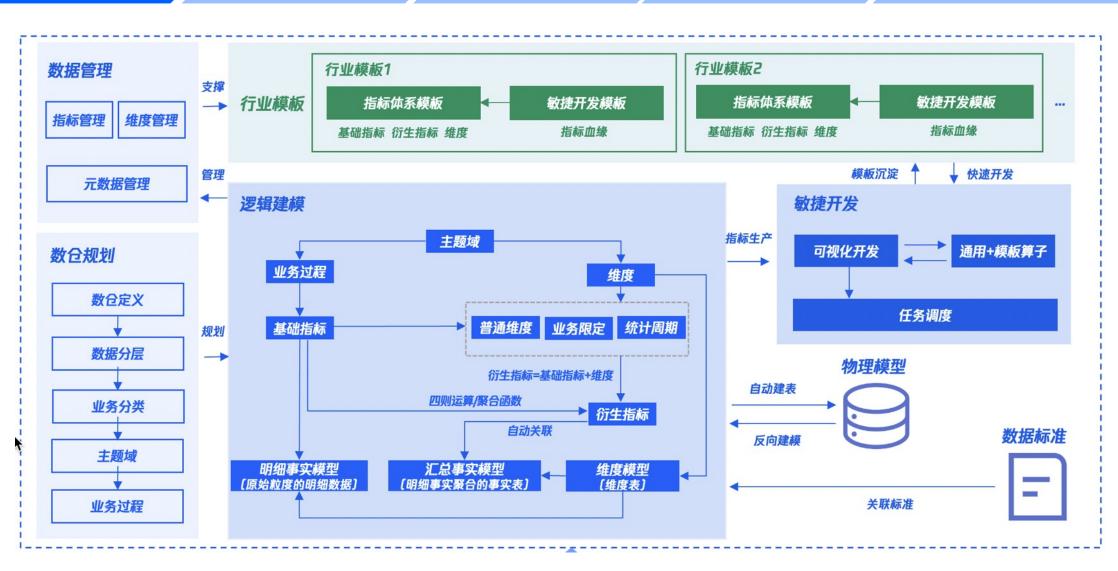


## WeData数据治理-规范工具





规范



## WeData数据治理-质量工具

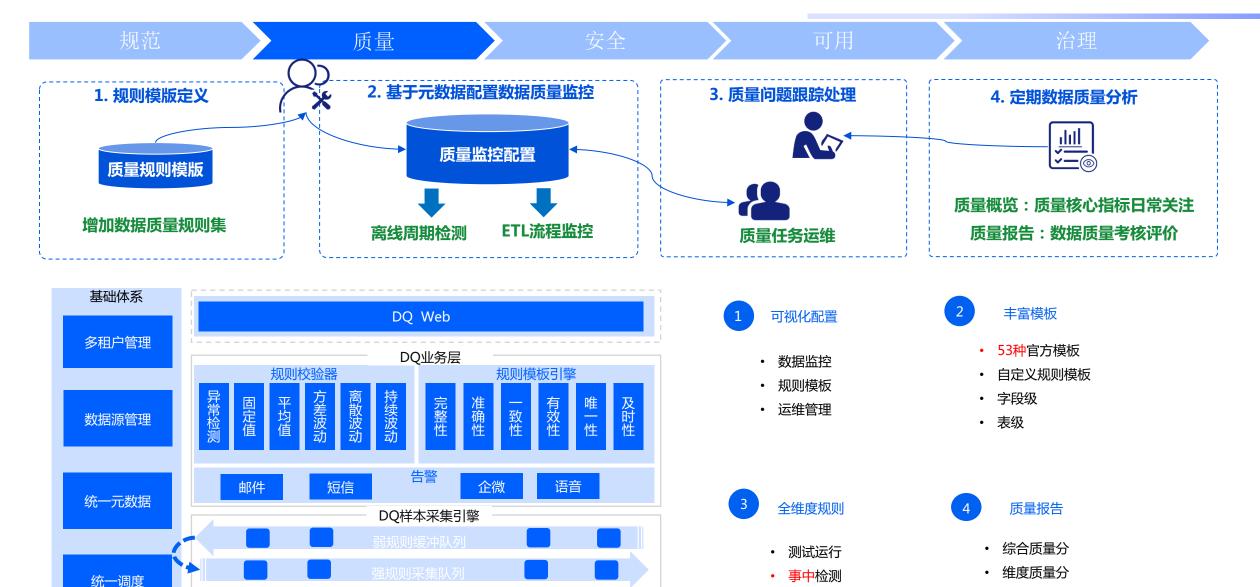
规则管理



• 表质量分明细

事后检测





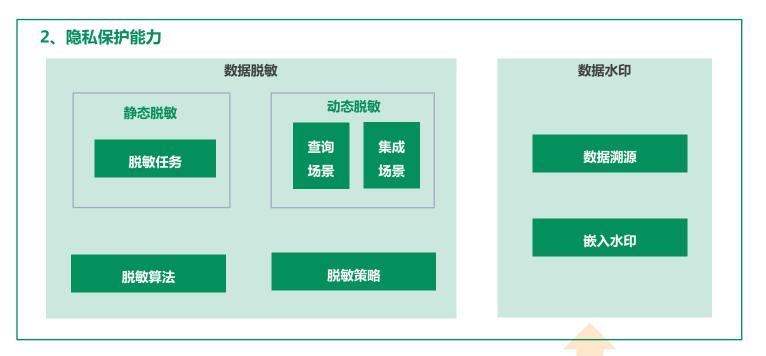
告警规则执行器

## WeData数据治理-安全工具





安全





#### 基于数据分类分级的安全能力



## WeData数据治理-元数据资产管理工具





可用



#### 元数据管理流程









1

元数据采集

2

血缘影响分析

3

元数据应用

4

元数据运营维护

- ・数据源配置
- ・元数据采集任务配置
- ・库表元数据采集
- ·ETL调度元数据采集
- ・业务元数据采集

- ・元数据检索
- ・系统、表与字段血缘关系分析
- ・系统、表与字段影响分析关系 分析
  - ・业务元数据分析

• ...

- ・数据资产展示
- ・数据字典管理
- ・业务指标管理
- 数据质量检查
- ・变更影响评估 • ...

- ・元数据采集监控
- · 元数据采集问题处理
- ・系统日常运维监控

• ...



#### 以数据应用为目的元数据资产工具

#### 数据发现

- ✓ 采集任务管理
- ✓ 采集实例运维
- ✓ 数据归属

#### 数据管理

- ✓ 库表管理
- ✓ 业务元数据管理

#### 数据地图

- ✓ 统计大盘
- ✓ 数据目录
- ✓ 血缘解析、数据温度、生产变 更.....
- ✓ 数据订阅

#### 元数据服务

- ✓ 元数据API
- ✓ 血缘API
- ✓ 数据统计API
- ✓ 元数据查询API

#### 基础能力支撑服务

统一元数据(MetaData)

统一安全(SecurityCenter)

统一调度(US)

## WeData数据治理-治理实践落地





可用 规范 质量 安全 治理





# 感谢聆听

