



第九届中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2018

猎聘网数据治理实践

DIG-满成见

DTCC
2018

2018.05.10 – 12 北京国际会议中心

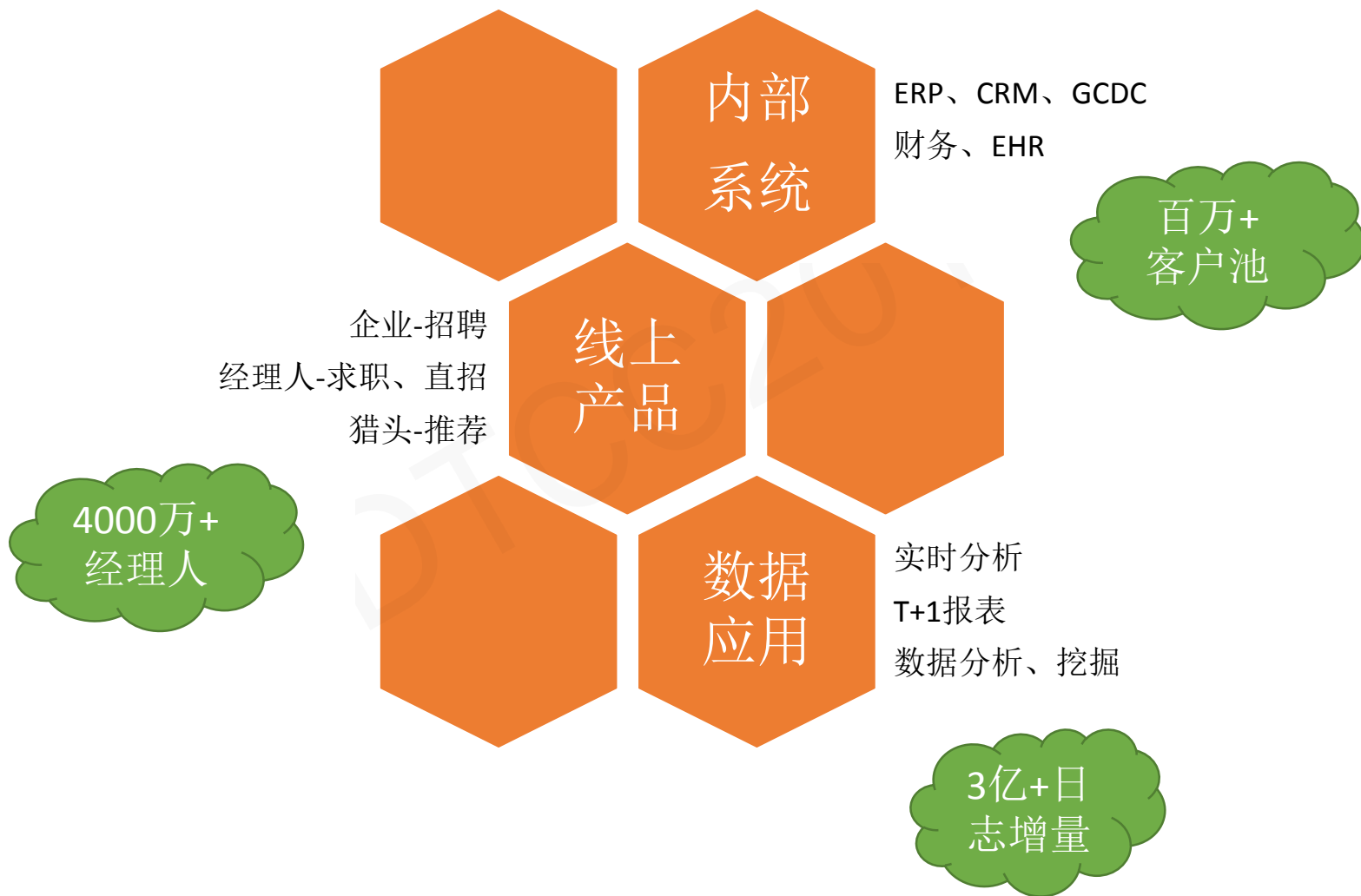


IT168.com

ChinaUnix

ITPUB

业务背景



业务困境

1. 没有专业的数仓团队
2. 没有模型规范管理，业务表到核心报表的一步到位
3. 近两千张表50%以上元数据缺失，命名规范不全
4. 数据质量完全依靠分析师个人判断，数据问题由下往上发现追溯
5. 缺乏数据上下游节点的合作规范和沟通流程

老板：报表指标怎么全部没数？

分析师：啊？我问下研发啊

研发：产品升级，数据迁库了

分析师：报告老板！报表停更三天，任务接口替换升级

部门A：报告老板，我们广告页UV已经达到百万级别了！！

部门B：全站DAU才几十万，你们从哪儿取的数据，去重了么？

研发A：今天新埋点刚上线，日志量预计会新增20%左右

分析师A：小C啊，你们产品一个月升级了4次，字段含义改来改去的，我十几个报表SQL，每次都要跟着改，公司盒饭都吃吐了

工程师C：知足吧，我这几天可都回回记得通知到你了

好的开端

1. CDO重视基础数据治理
2. 大数据团队的接口层设计完善，完整的保留了90%以上的核心业务数据每天的拍照
3. 任务调度系统能够记录任务与任务之间的血缘关系
4. 数据在线查询平台提供了原始的元数据管理平台



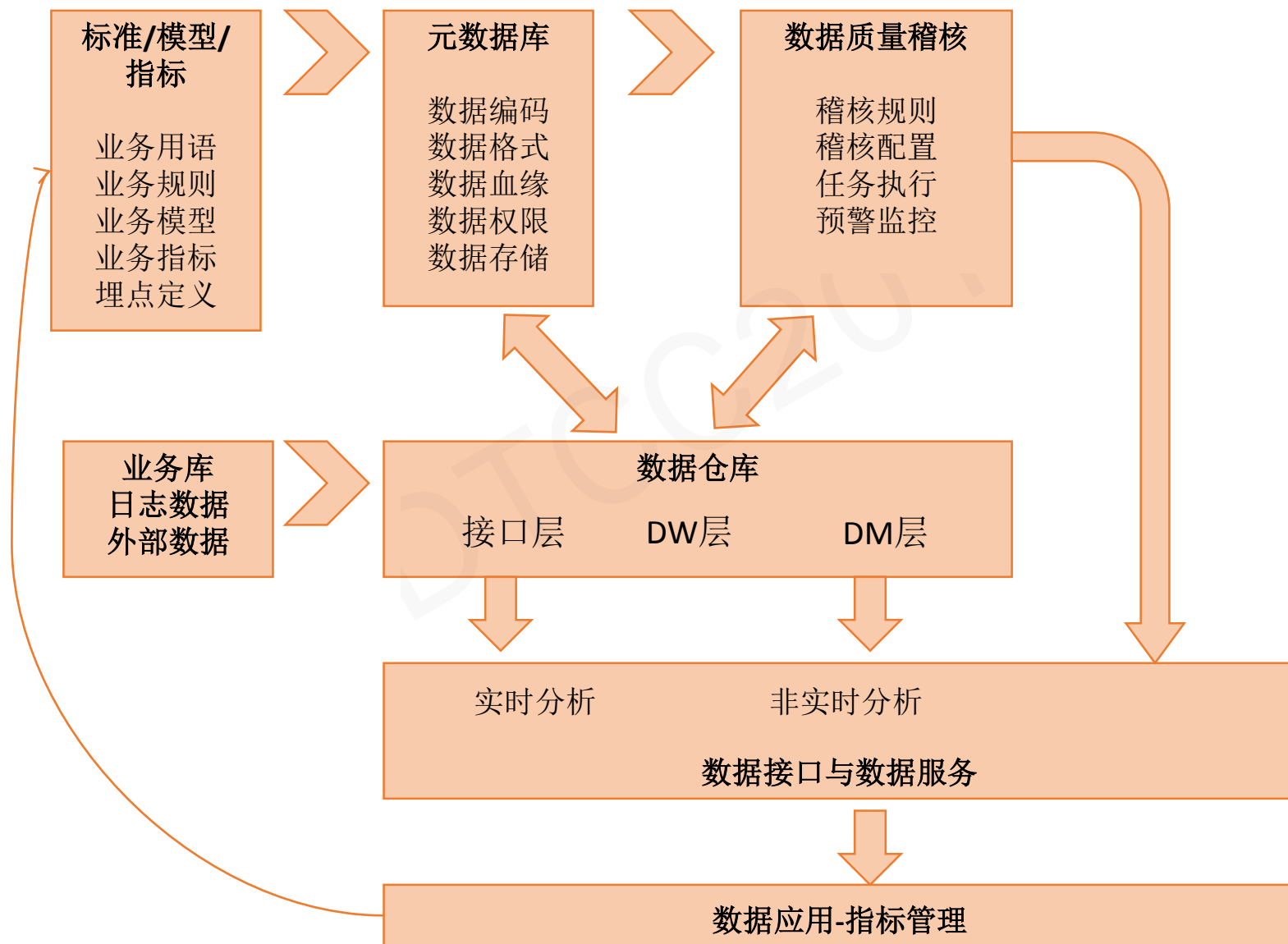


模型架构

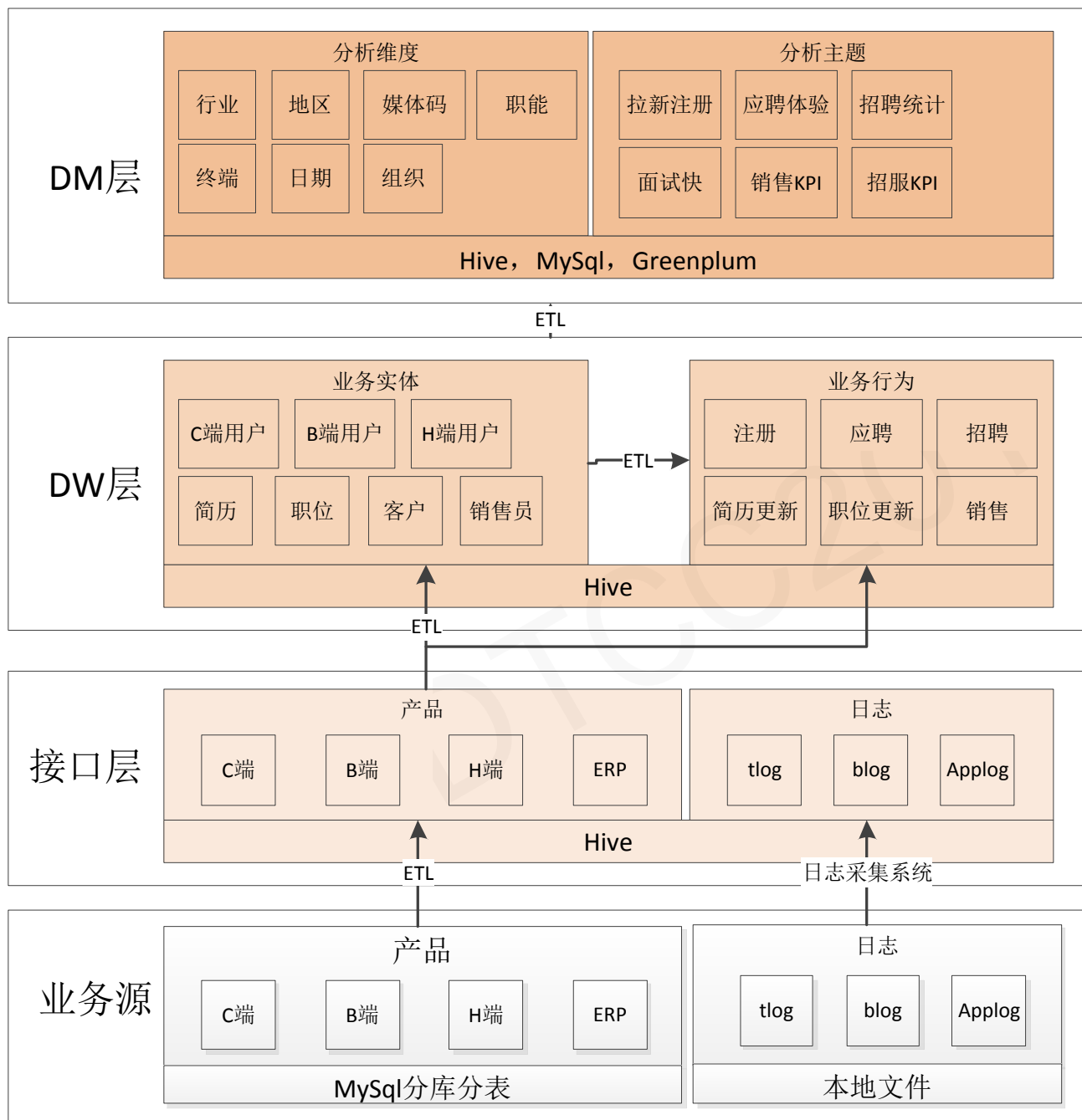
平台技术

流程规范

数据治理流程



数据仓库架构



数仓分层管理规范

接口

- 保留历史：业务表全量拍照，按天存储增量分区，全量和增量整合去重保持与业务一致，开放给用户
- 数据规整：分库分表整合，生成唯一主键，编码转换，敏感字段过滤，数据格式统一

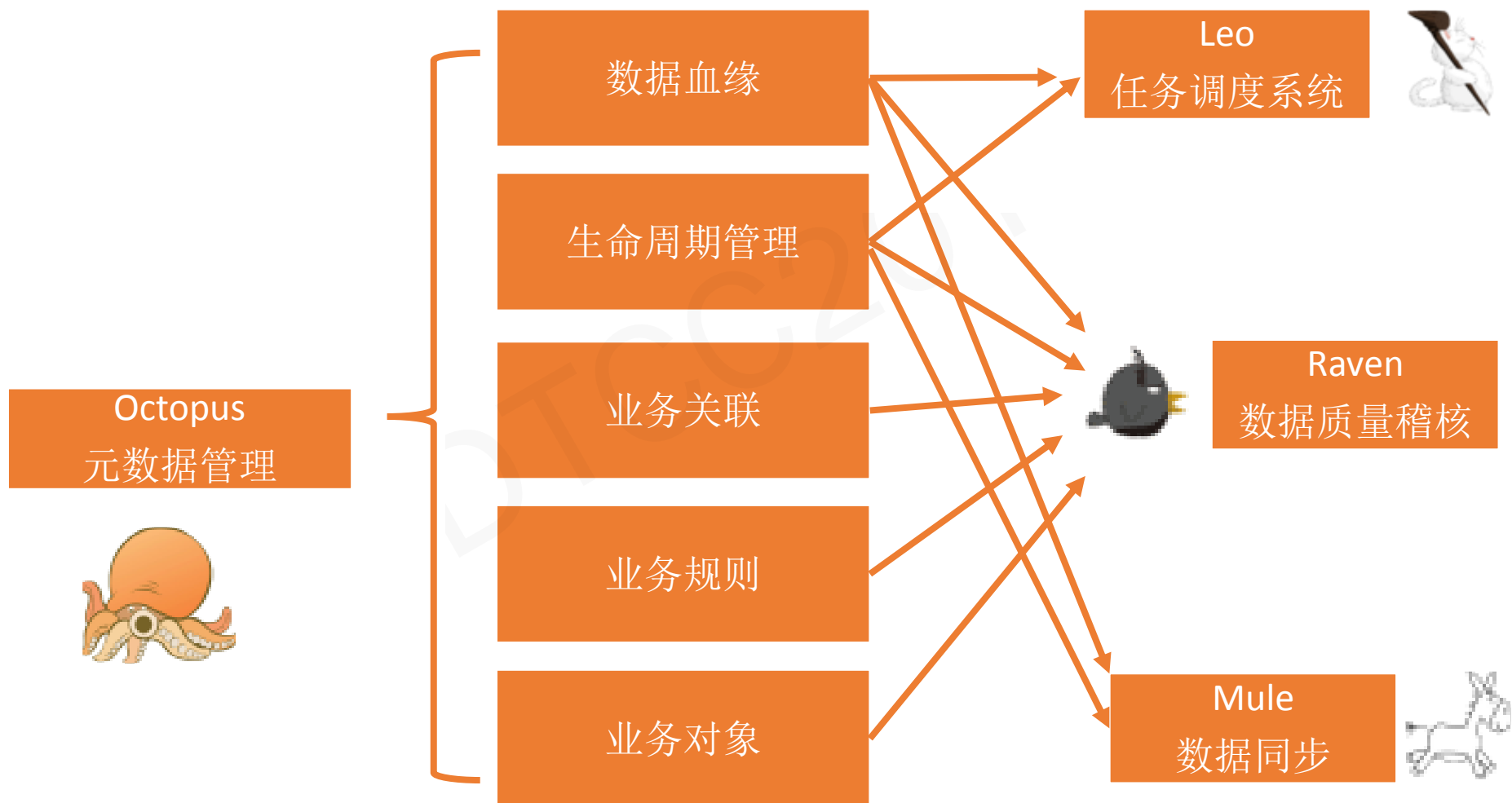
DW

- 划分业务主题，关联、整合业务实体信息，汇总统计实体行为指标，宽表设计提高查询、分析效率
- 关键实体信息周期拍照，支持业务发生时上下文分析
- 统一数据应用出口，屏蔽业务源变更影响

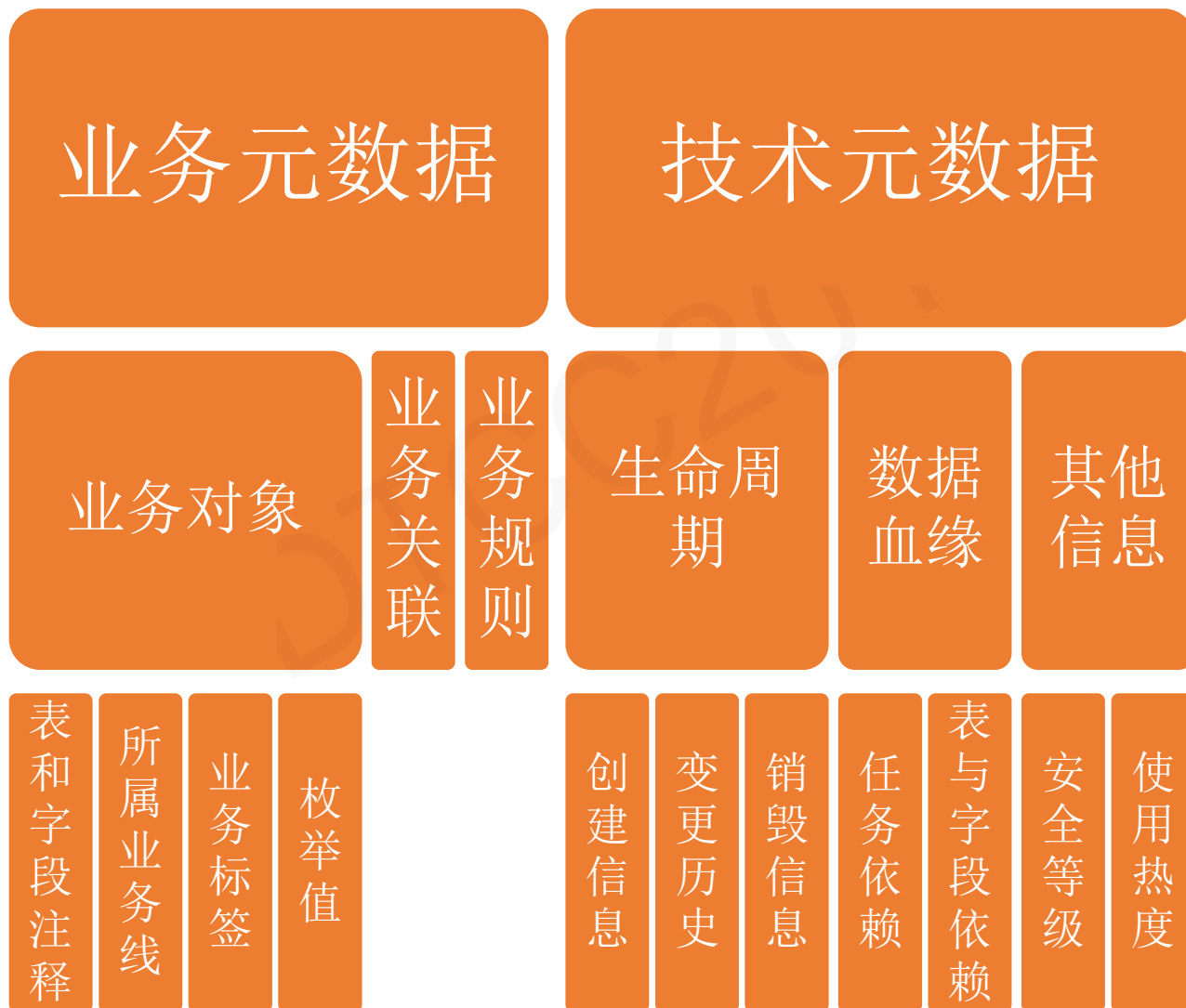
DM

- 决策层核心报表与部门级集市报表
- **Hive**保留原始加工数据，**Mysql**支持T+1报表，**Greenplum**支持10分钟以内的实时业务分析

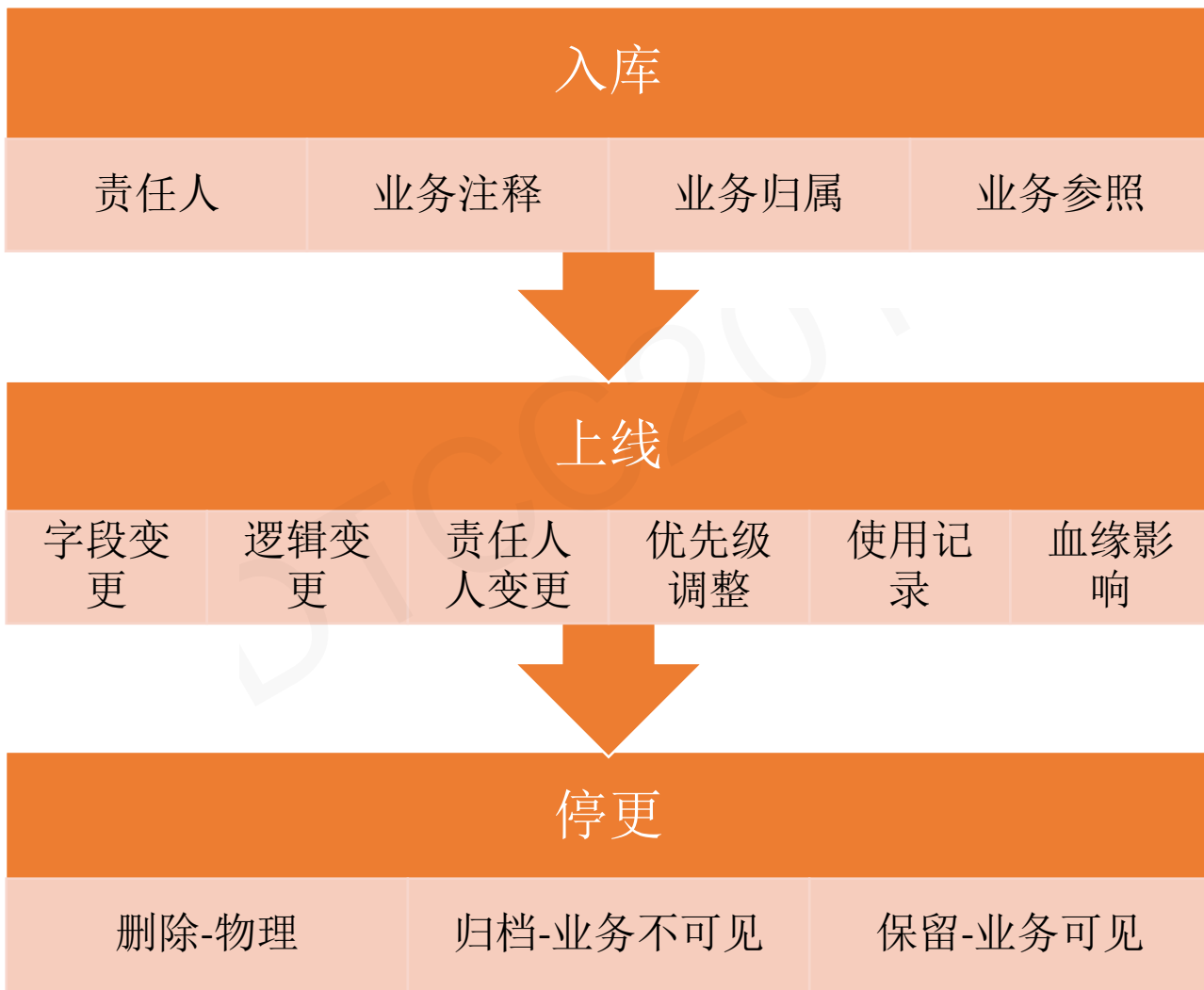
技术平台



元数据管理



数据生命周期管理



数据血缘影响

血缘关系

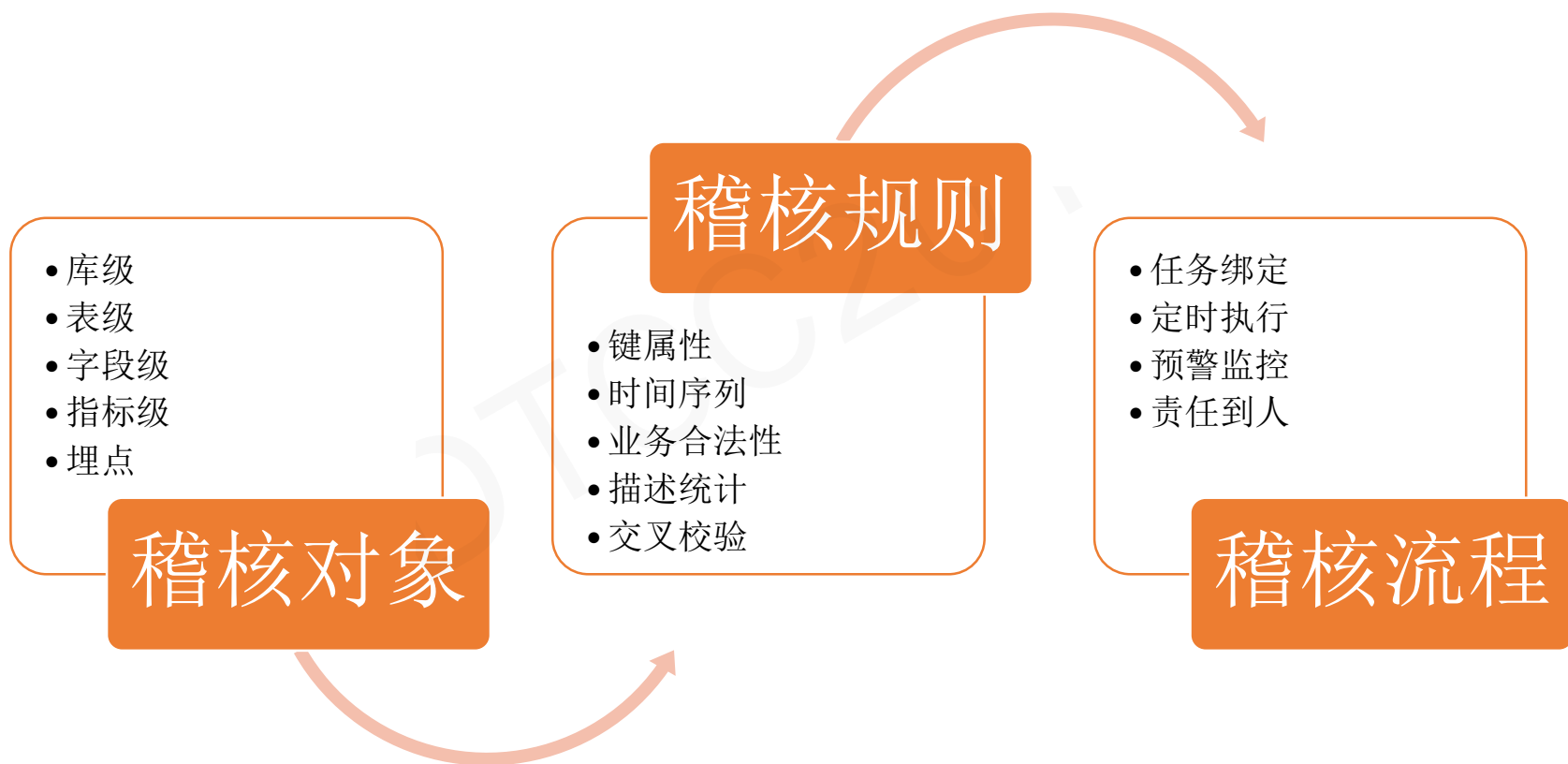
- 表与ETL任务
- 表与数据表
- 表与数据表的字段
- 表与埋点事件
- 表与UDF函数
- 表与指标
- 表与报表



数据血缘的应用

- ETL的DAG图
- 任务优先级制定
- 销毁上游对象约束
- 质量稽核影响性
- 打通各平台工具的脉络

数据质量稽核



数据质量稽核规则与流程



规则配置

结果比对



数据质量稽核应用

洞察业务变动

- 数据量为空
- 同环比波动
- 参照异常
- 分布变更

探查质量问题

- 唯一性
- 参照异常
- 主从一致
- 异常值

数据质量报告

- 可靠性
- 可用性
- 业务评价
- 治理成效标准

数据安全

安全标准

- 个人隐私
- 业务敏感
- 财务数据
- 部门隔离
- 等级区分

数据粒度

- schema
- 表级
- 字段级
- 任务级
- 接口级

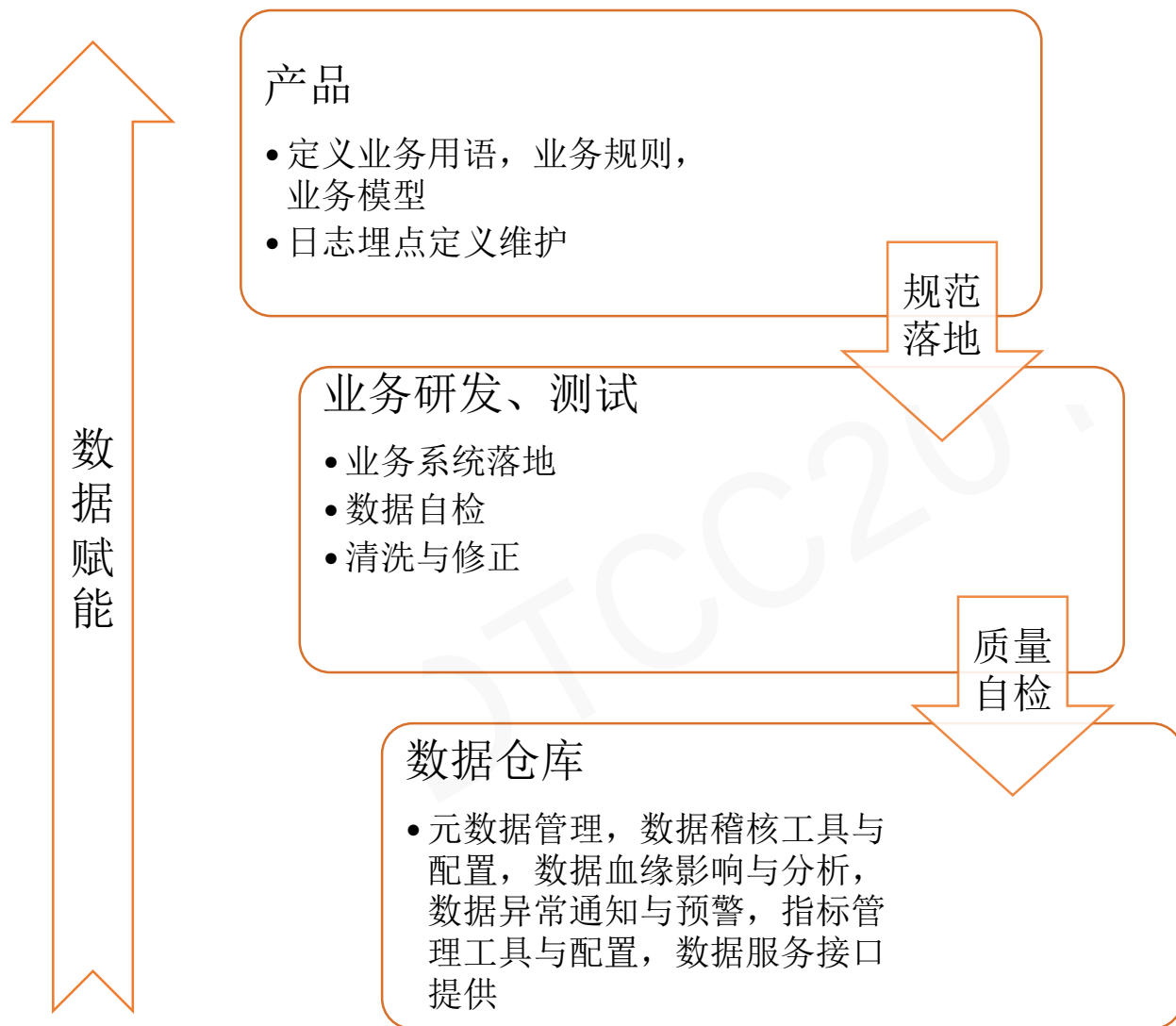
用户行为

- 查询
- 下载
- 邮件

审批流程

- 自动解析
- 系统提醒
- 人工审核

数据治理参与角色



成果



1. 定时监控上游数据结构与质量问题，发现问题当天跟进处理，保障90%以上的核心报表与数据应用的稳定与可靠
2. 模型分层，统一了下游引用出口，上游系统升级改造，影响截止到DW层，节省分析师90%以上的资源投入
3. 元数据完善95%以上，构建数仓知识体系与数据标准，业务与数据学习效率极大提升

挑战

1. 数据治理中心的推广：将数据治理推进到研发一线，从源头提高业务数据质量。如何将数据质量提升到全司共识？
2. 日志埋点数据的治理：定义、研发、测试、上线整个流程中涉及序列人员众多，如何规范？如何避免人为影响？

THANKS





讲师申请

联系电话（微信号）：18612470168

关注“ITPUB”更多
技术干货等你来拿~

与百度外卖、京东、魅族等先后合作系列分享活动



让学习更简单

微学堂是以ChinaUnix、ITPUB所组建的微信群为载体，定期邀请嘉宾对热点话题、技术难题、新产品发布等进行移动端的在线直播活动。

截至目前，累计举办活动期数60+，参与人次40000+。

ITPUB学院

ITPUB学院是盛拓传媒IT168企业事业部（ITPUB）旗下
企业级在线学习咨询平台
历经18年技术社区平台发展
汇聚5000万技术用户
紧随企业一线IT技术需求
打造全方式技术培训与技术咨询服务
提供包括企业应用方案培训咨询（包括企业内训）
个人实战技能培训（包括认证培训）
在内的全方位IT技术培训咨询服务

ITPUB学院讲师均来自于企业
一些工程师、架构师、技术经理和CTO
大会演讲专家1800+
社区版主和博客专家500+

培训特色

无限次免费播放
随时随地在线观看
碎片化时间集中学习
聚焦知识点详细解读
讲师在线答疑
强大的技术人脉圈

八大课程体系

基础架构设计与建设
大数据平台
应用架构设计与开发
系统运维与数据库
传统企业数字化转型
人工智能
区块链
移动开发与SEO



联系我们

联系人：黄老师
电话：010-59127187
邮箱：edu@itpub.net
网址：edu.itpub.net
培训微信号：18500940168