教育大数据治理实践

科大讯飞-王龙生

ITPUB.NET

个人简介：

科大讯飞教育事业群大数据部门经理

资深软件架构师、资深数据架构师、业务专家

从事教育行业8年，负责班班通、智慧课堂、智学网等核心产品研发，带领团队从0构建教育大数据中台

ITPUB.NET

CONTENTS 大纲

- 业务背景
- 数据治理框架
- 数据问题及治理方案
- 总结

ITPUB.NET

教务
课堂
作业
考试
家校通
...
业务系统数据
网络日志
机器数据
硬件平台

ITPUB.NET

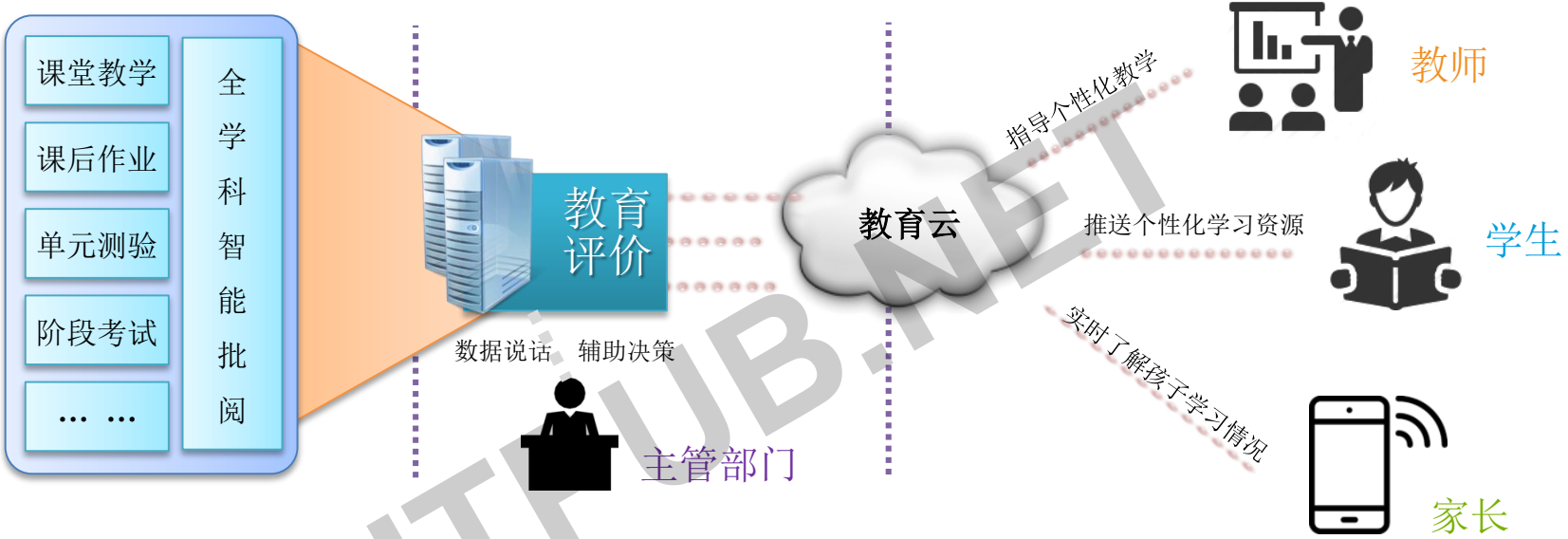
产品多，功能重复建设

渐进性发展，业务竖井

互联网业务，版本迭代快

智学网 zhixue.com

ITPUB.NET



1 教学过程数据采集与汇聚

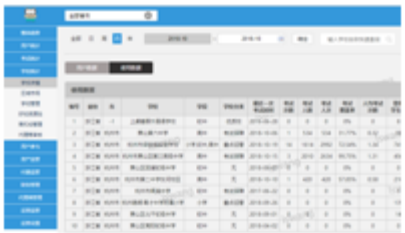
2 教育大数据自动分析与评价

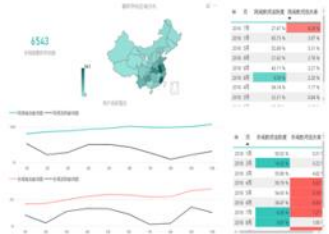
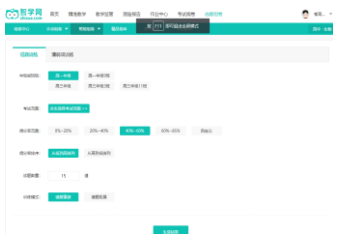
3 应用大数据实现个性化教学

人工智能助力教育，因材施教成就梦想

ITPUB.NET

数据集市开发 → 大数据交付项目 → 大数据DI辅助产品 → 大数据BI辅助决策 → 大数据AI推荐



PUB.NET

数据缺失

信息共享困难

不准确

问题定位难

口径不一致

问题解决周期长

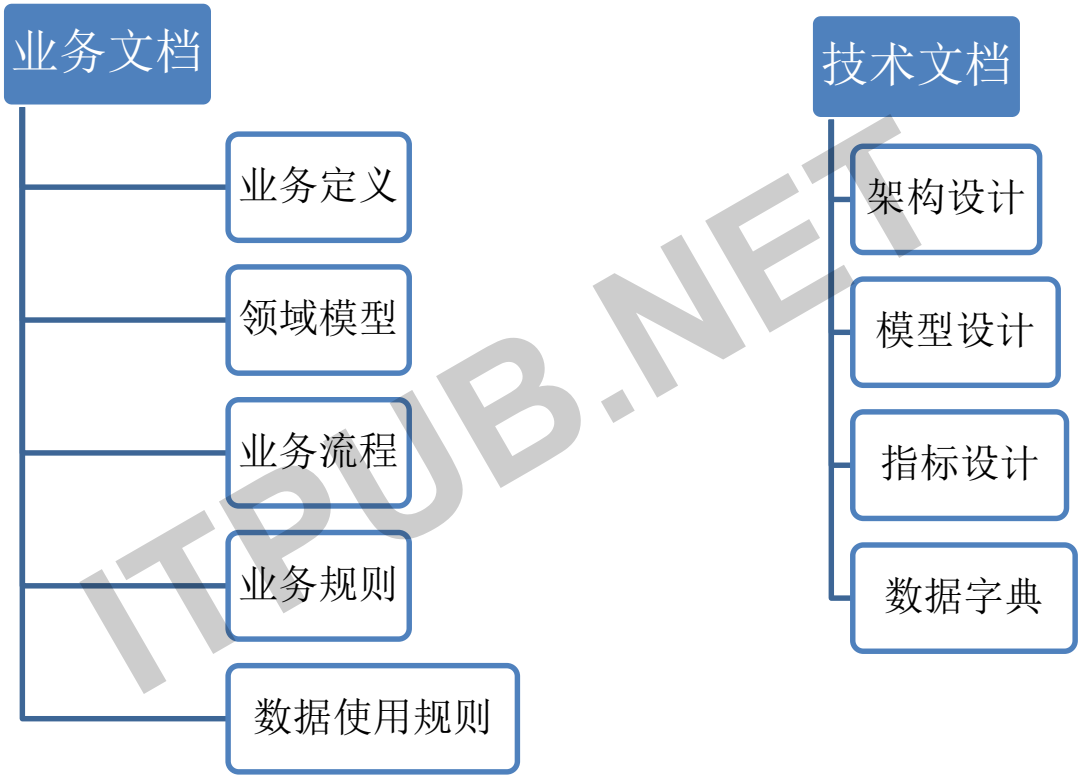
……

ITPUB.NET

数据治理框架

ITPUB.NET

文档管理



ITPUB.NET

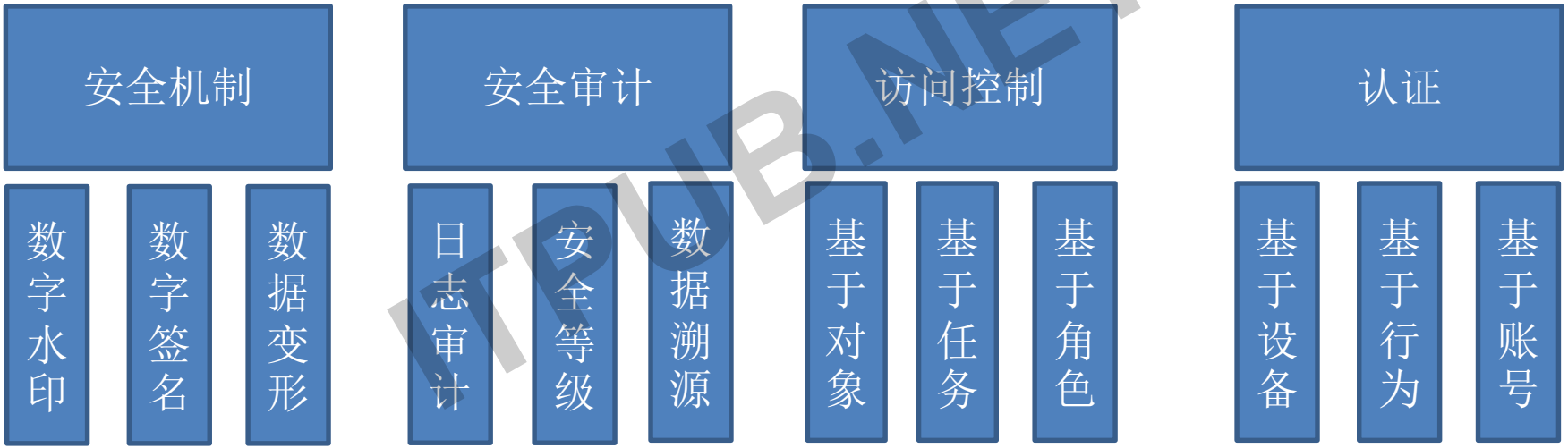
在线文档管理

数据安全管理

ITPUB.NET

技术是关键、策略是核心、管理是保证

数据安全管理



ITPUB.NET

元数据管理

业务元数据

- ✓ 业务名称
- ✓ 业务定义
- ✓ 业务描述
- ✓ 业务规则
- ✓ 业务流程
- ✓ 领域模型

技术元数据

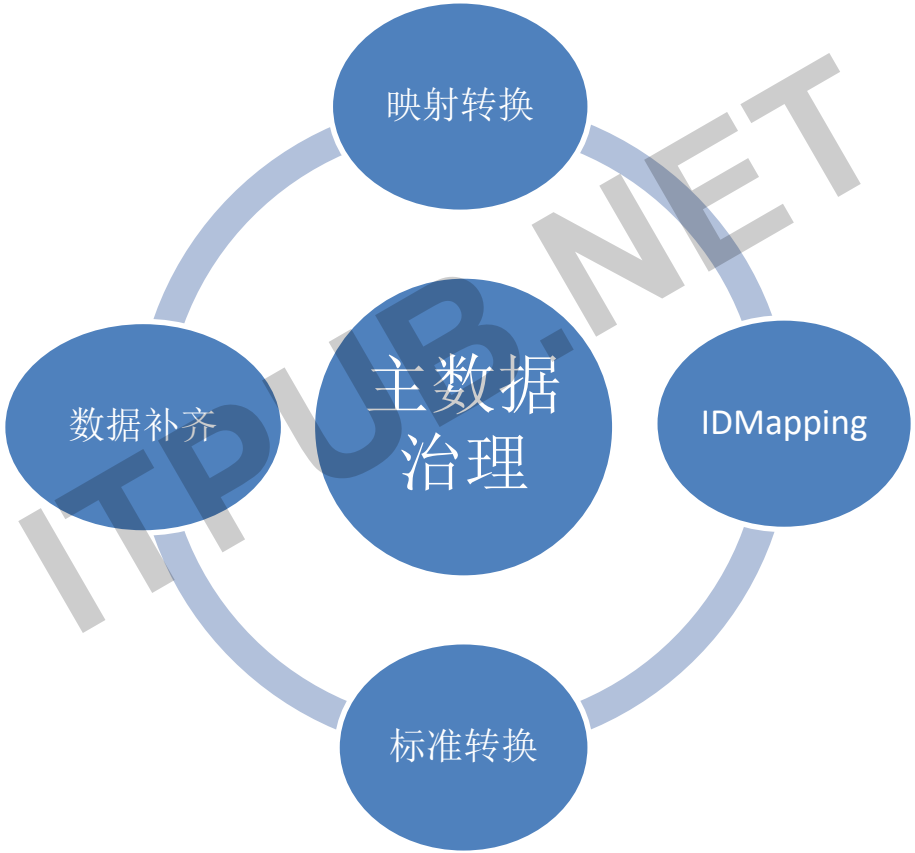
- ✓ 库表结构
- ✓ 字段定义
- ✓ 血缘关系
- ✓ 变更历史
- ✓ 使用热度
- ✓ 保留时长

元数据应用

- 数据地图
- 血缘关系
- 使用热度
- 任务依赖
- 安全等级
- 影响分析
- 表标签管理
- 变更历史

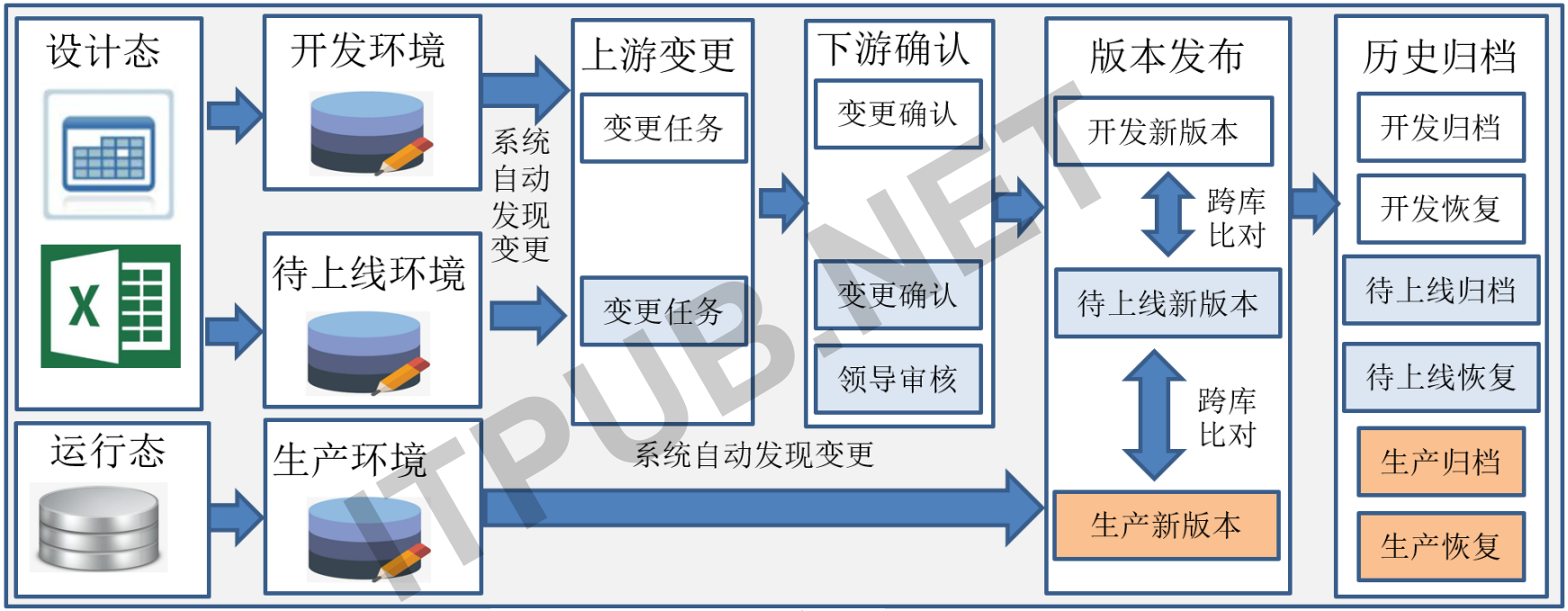
ITPUB.NET

主数据治理



ITPUB.NET

变更管理



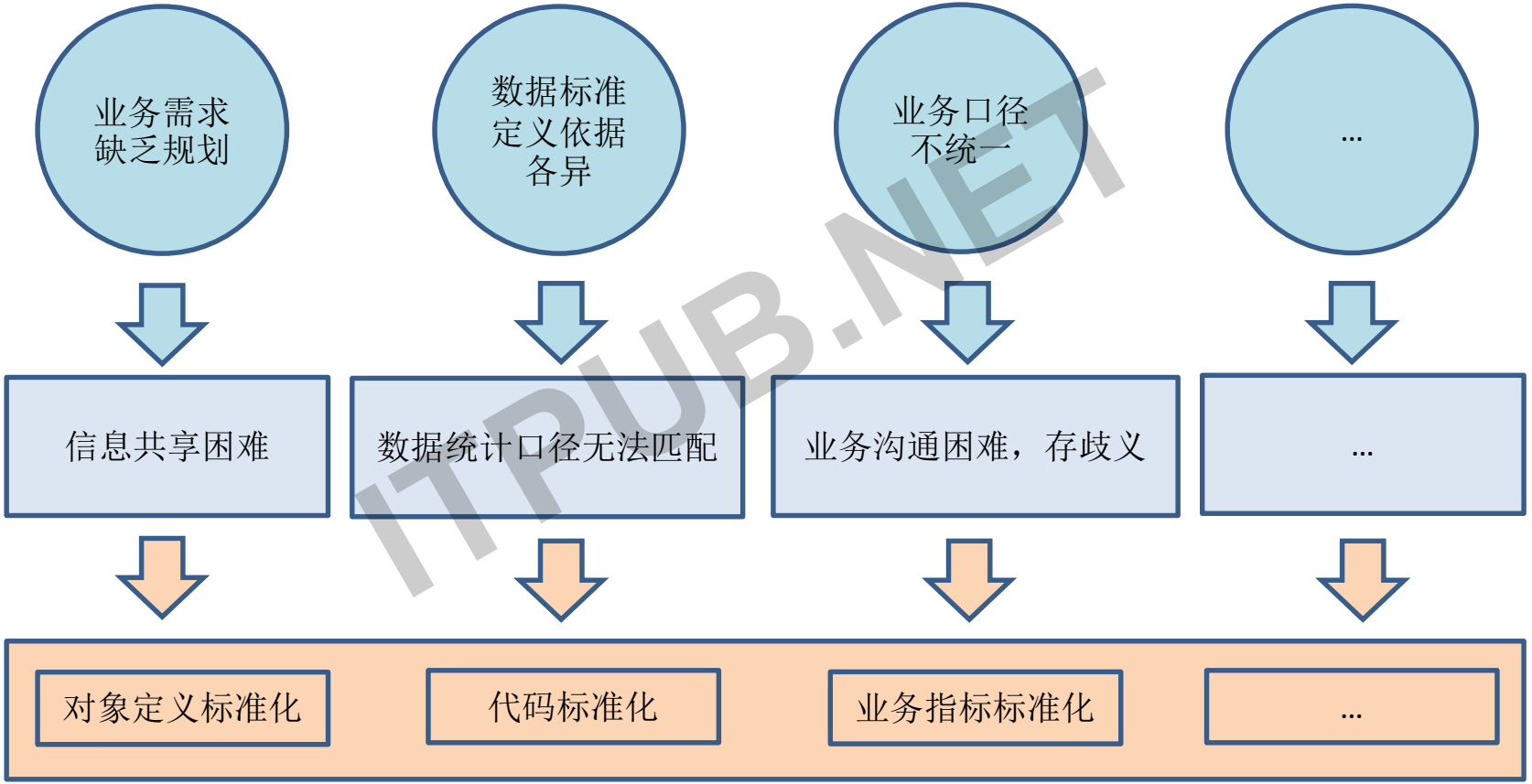
ITPUB.NET

时效性
- 定时采集
- 获取变更

全流程
- 自动通知
- 事中监控

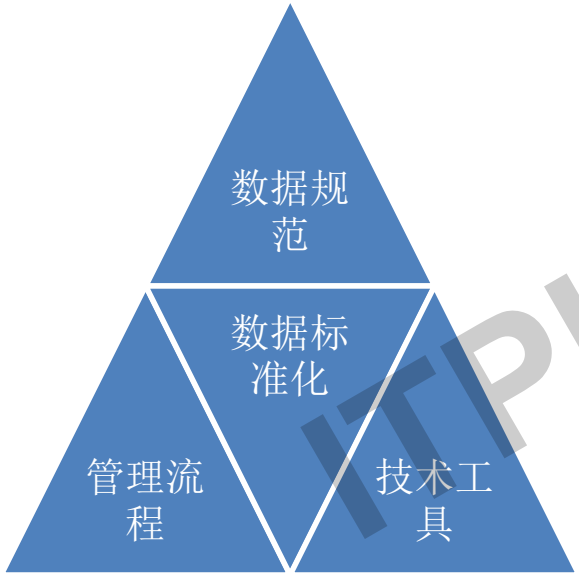
一致性
- 跨库比对
- 事后检测

标准问题



ITPUB.NET

标准管理



- ✓ 数据代码类标准
- ✓ 词库编码工具
- ✓ 建模工具
- ✓ 指标管理系统
- ✓ 标准检查工具

ITPUB.NET

数据质量管理

协调与组织业务策略保持一致

检测数据质量

建立数据质量管理角色和责任

维护、管理及应用数据

质量管理的支撑工具

ITPUB.NET

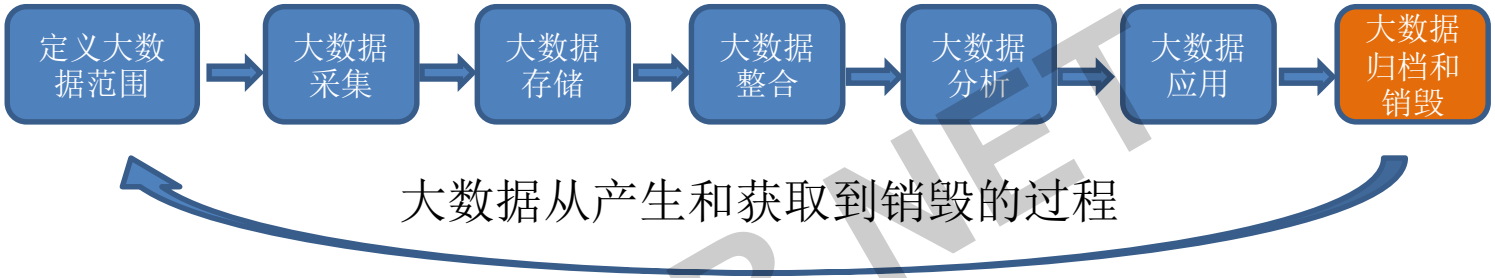
数据质量管理

质量分析：提供不同值的频度分布，对每个字段类型和用途的洞察分析

- ✓ 数据量波动
- ✓ 枚举异常
- ✓ 主键不唯一
- ✓ 业务规则监控
- ✓ 数据质量报告

ITPUB.NET

数据生命周期管理



ITPUB.NET

- ◆ 分析
- ◆ 归档 → • 可选择性恢复
- ◆ 销毁 → • 严格审批流程

数据生命周期管理

分类	热数据	温数据	冷数据
数据价值密度	高	中	低
数据使用频度	高	中	低
数使用方式	静态报表	数据分析	数据检索
数据使用目的	支撑进行决策	分析有意义数据	寻找有意义的数据
数据存储量	低	中	高
数据使用工具	可视化展现工具	可视化分析工具	编程语言和技术工具
使用使用者	决策者	业务分析者	数据专家

数据热度
- 热数据
- 温数据
- 冷数据

- 高性能、高并发、高可用、高可靠
- 高性能、高可靠
- 低成本、低并发、大容量、可扩展

不同的热度采用不同的存储和备份策略

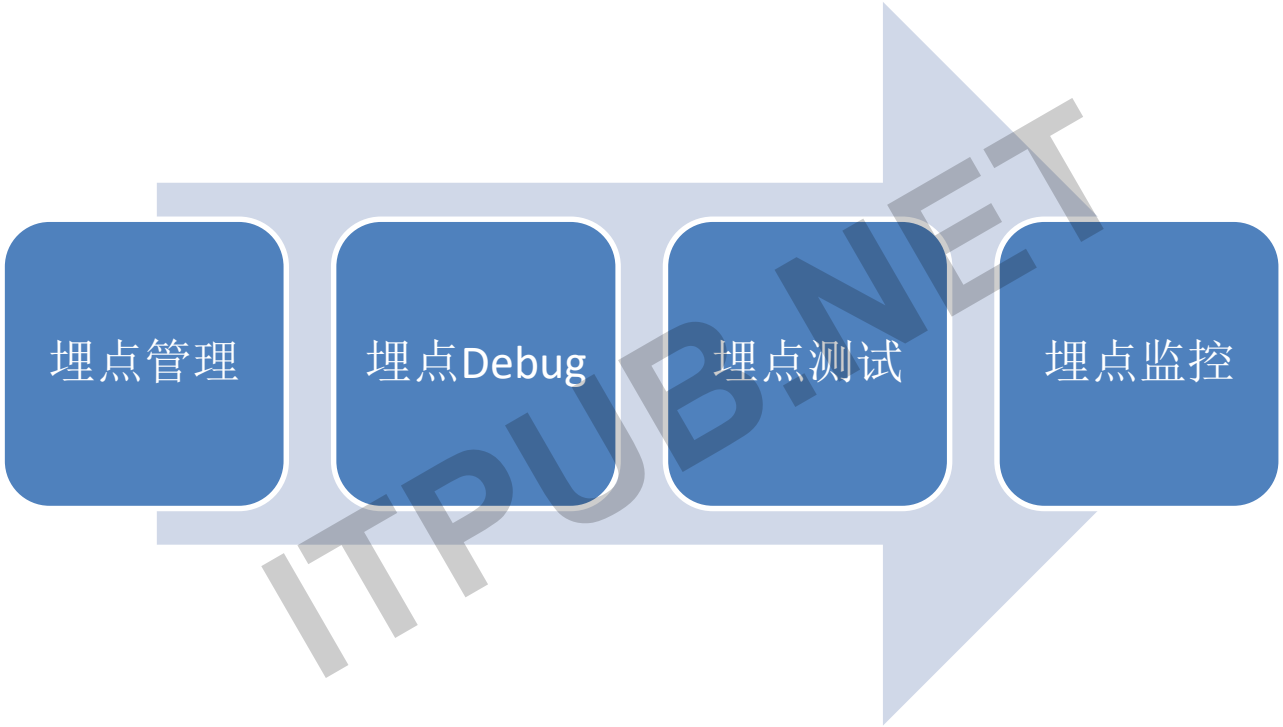
ITPUB.NET

制度管理

- 质量保障规范
  - ✓ 质量巡检制度
  - ✓ 线上变更制度
  - ✓ 模型复盘制度
  - ✓ 运维值班制度
  - ✓ 数据安全审计
  - ✓ 指标设计规范
- 流程管理规范
  - ✓ 上线规范
  - ✓ 采集规范
  - ✓ 模型变更流程
  - ✓ 生命周期管理规范
  - ✓ 数据导出规范
  - ✓ 仓库问题处理规范

ITPUB.NET

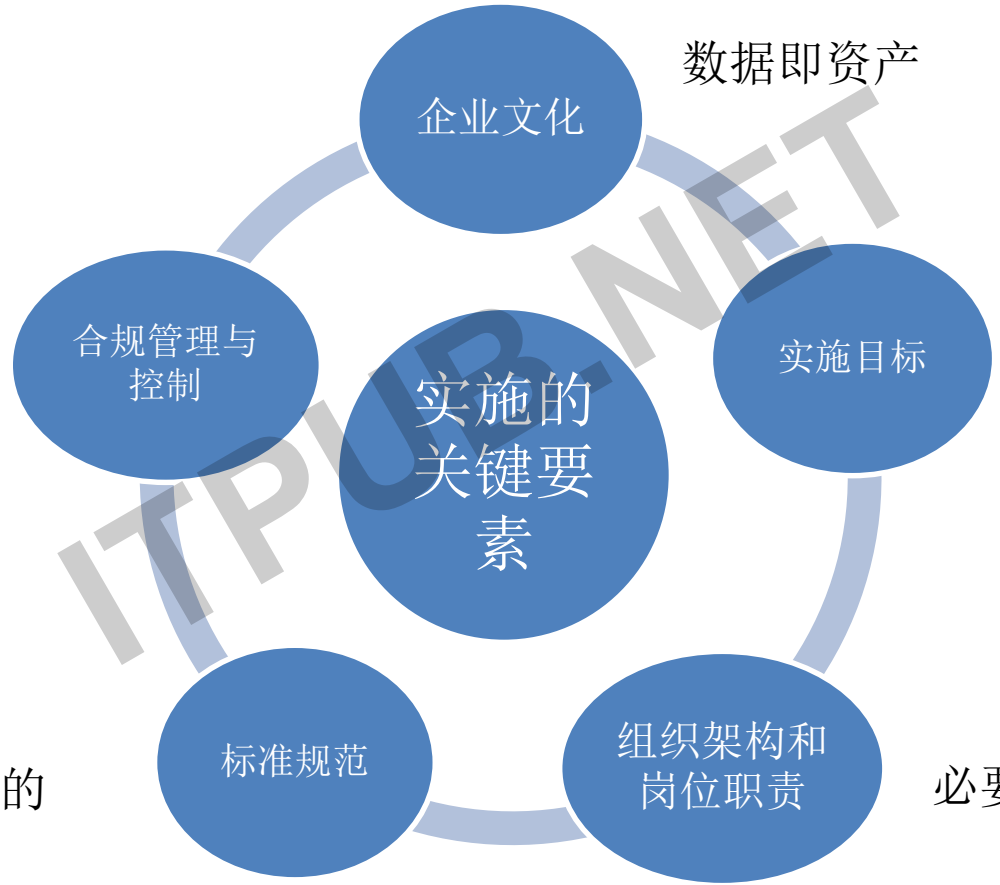
日志数据治理



ITPUB.NET

技术可以降低成本，核心是业务驱动

数据实施框架



数据即资产

依据业务发展需求，设立合理的阶段性实施目标；

必要保障

数据治理治理的前提条件

流程化，工具化，采用科学手段高效实施大数据治理

ITPUB.NET

ITPUB.NET