



2019

05

08-10

北京新云南皇冠假日酒店

数据风云 十年变迁

DTCC

第十届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2019



+

○

○

○

DTCC 2019

第十届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2019

大数据时代的数据治理

石竹软件总经理

薛勇

CONTENTS
目录

01

大数据时代数据的特点

02

大数据时代的数据治理

03

典型案例

企业信息系统演进的时代

OLTP时代

- 企业主要的IT系统
ERP\进销存
MRPII\SCM
OA系统
HR系统
....
- 主要特点
业务的电子化
数据的电子化

OLAP时代

- 企业主要的IT系统
DW/BI系统
统一客户管理系统
报表系统
CRM系统
...
- 主要特点
内部数据统一
单向数据应用
强调数据的技术

BigData时代

- 企业主要的IT系统
大数据平台
数据湖
数据中台
....
- 主要特点
内外部数据统一
双向数据应用
数据既是技术也是资源

大数据时代的数据特点

- 多源
 - 内部系统、外部系统
 - 文件系统、RDBMS、Hadoop、实时流式
- 异构
 - 结构化、半结构化、非结构化
 - 数字、图象、声音、视频
- 海量
 - TB、PB、ZB
- 角色变换
 - 由附属的信息变为资源，本身可以盈利



大数据时代企业面临的数据困境

技术人员的困境

- 数据变化更难以跟踪；上游系统发生变更，下游系统不知道；
- 多业务系统之间数据定义不统一、格式不统一、统计口径不统一，形成多个信息孤岛；
- 数据质量达不到要求，影响数据集成和整合；
- 数据来源混乱，版本流失，数据权责不清。

业务人员的困境

- 理解困难：含义相同的数值在不同的统计报表中，差异巨大，而且由转换轨迹模糊，无法排查数值偏离原因；
- 数据使用效果有限：多个来源数据的质量不高，在低质量数据基础上进行数据挖掘与分析，结论的有效性和可用性打折扣。

企业管理的困境

- 很难摸清数据资产的家底：系统过多、数据过杂、无法有效的统计和获知数据资产的真实情况。
- 知识传承与业务衔接低效、开发、运维人员的流动，造成系统运转不畅、知识的丢失
- 数据作为资产难以管控？数据资产不能有效的保值增值

大数据时代更需要数据治理

4

将数据作为资产进行管控

越来越多的企业将数据视作一种资产，对这类特殊的资产如何进行管理，如何进行价值评估也是数据治理重要的组成部分。

1

形成跨系统的数据地图

企业为了更好管理自身多源异构的庞大数据，需要通过元数据建立大数据目录，形成跨系统的全面数据地图。

数据治理

2

高质量的数据才能产生高质量的结果

企业通过对数据的分析挖掘发现新的问题和机遇，只有高质量的数据才能保证结果的正确性。多源数据质量驳杂，更需要通过治理提高数据质量。

3

升级数据安全、生命周期的管理

大数据时代，数据的流动性和时效性进一步增强，由此对数据安全、数据生命周期的管控要求也进一步增强。



CONTENTS

目录

01

大数据时代数据的特点

02

大数据时代的数据治理



数据治理咨询

数据治理产品

数据治理实施

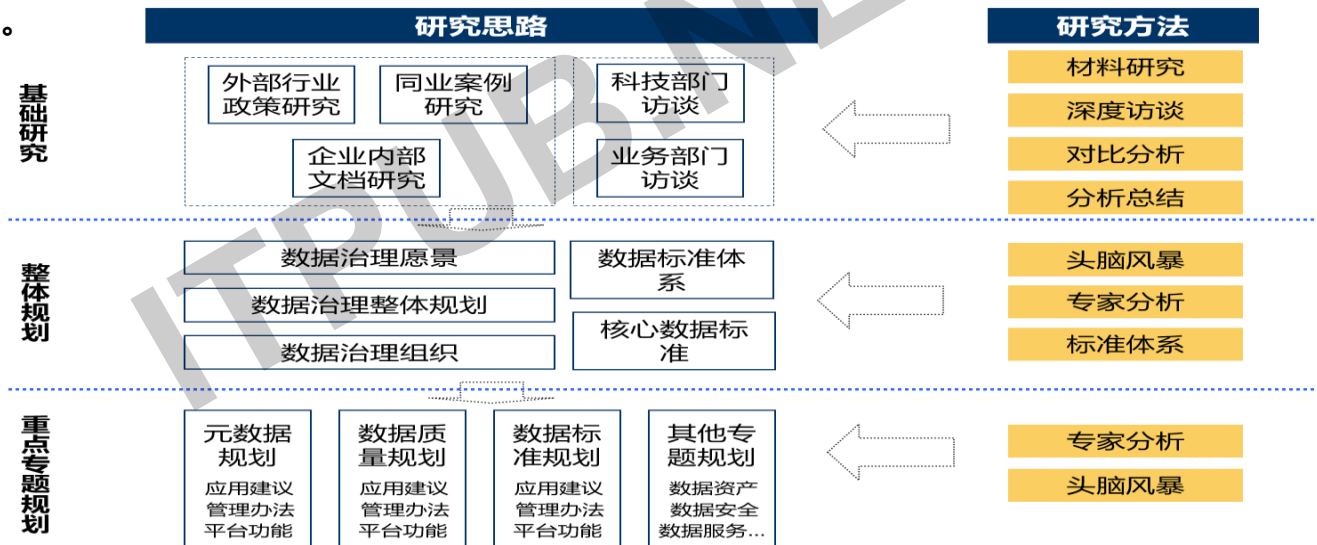
03

典型案例



数据治理是自上而下的系统工程

- 数据治理是全企业的、长期的系统工程，企业需要将数据治理作为企业战略的一部分，而不是某个部门某一时期的工作。
- 数据治理之初，需要对企业内外部进行调研，制定相应的目标、组织和规范、要划分阶段并明确各阶段重点工作。



在企业中建立专门的数据治理机构

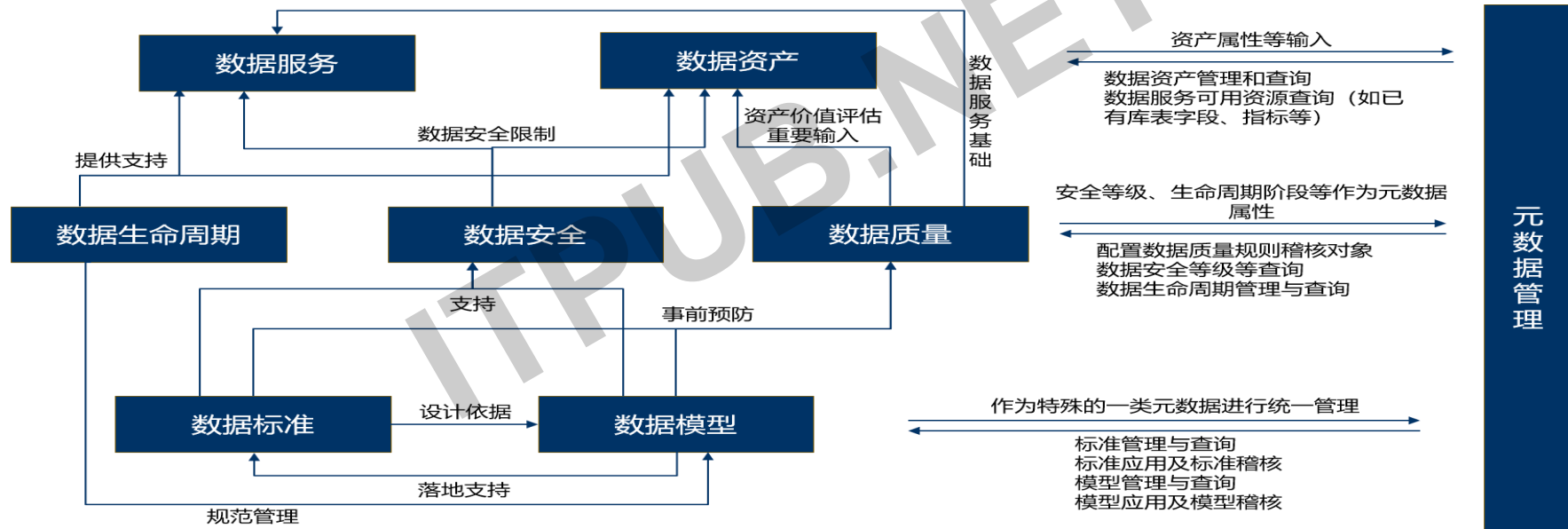
- 数据治理机构的建立是企业数据治理必不可少的一个环节，通过数据治理机构保障数据治理工作落实到各个部门和系统。
- 数据治理机构还肩负着制定数据治理相关规章制度、流程、考核方式等工作。



示例

分析明确企业数据治理实施专题

- 数据治理包含多方面的工作，数据治理机构需要根据整体目标和实际情况，确定自身数据治理主要进行哪些方面的提升。



CONTENTS
目录

01

大数据时代数据的特点

02

大数据时代的数据治理

数据治理咨询



数据治理产品

数据治理实施

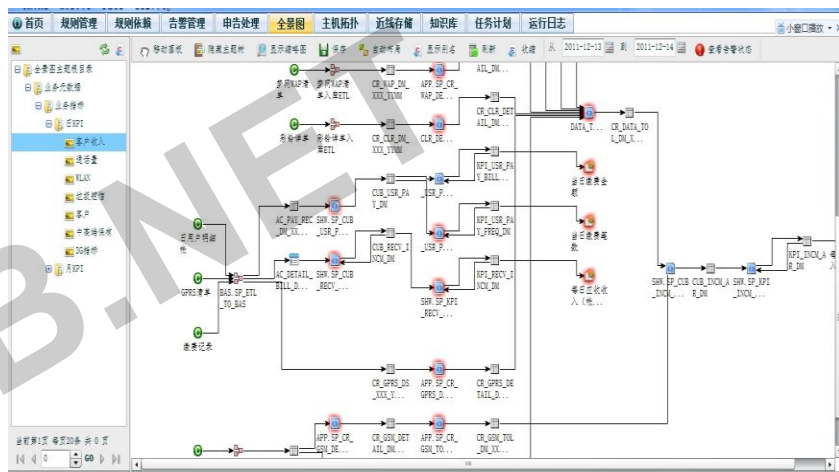
03

典型案例

数据治理的主要技术与产品



元数据：企业数据治理的基础

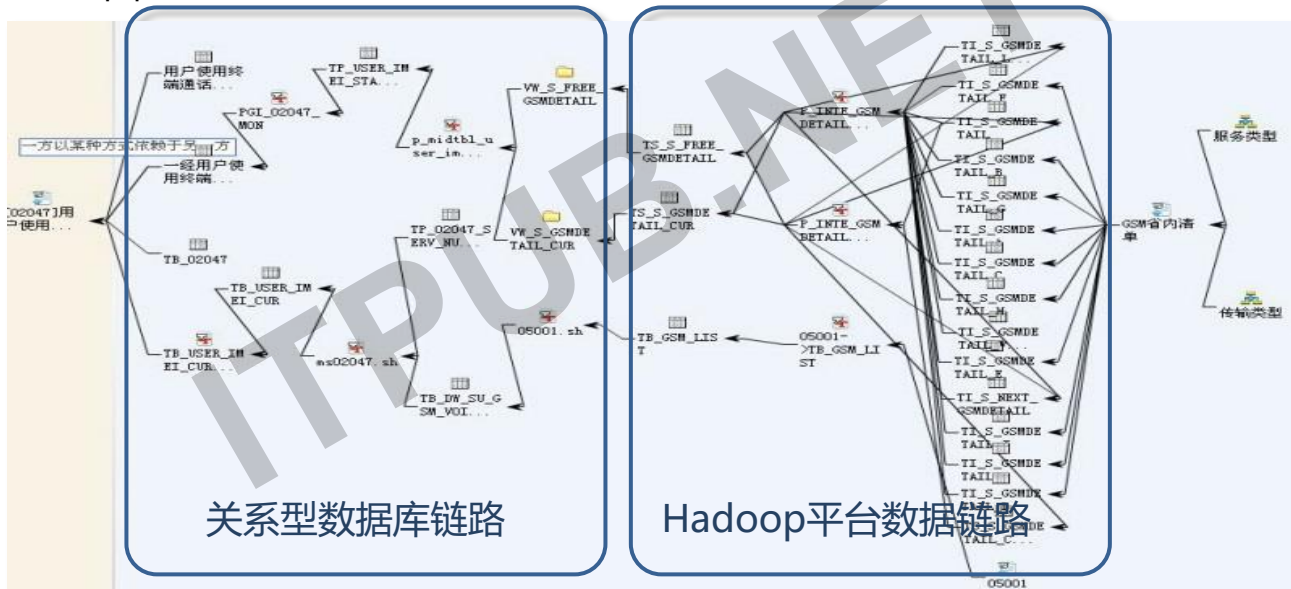


建立企业全局性的数据字典和数据地图功能，对整个系统数据定义和数据流向一目了然。



元数据：形成全面的数据地图

- 元数据系统可以支持关系型、MPP、分布式等多种类型数据库和相关ETL作业、BI工具，形成跨系统的完整地图。



元数据：对企业的数据资产情况一目了然



数据质量：对企业各环节数据进行检核

- 企业数据产生—加工—分析的各个环节都可能产生数据质量问题，通过创建数据质量检核规则，对企业数据的准确性、一致性等进行检核。

分类导航							Q 请输入搜索关键字 搜索	
规则分类： 所有 完整性 有效性 精确性 一致性 唯一性 时效性 真实性								
项目： 所有 属性3取值范围检查 属性5符合下拉列表检查 属性4正则表达式检查 属性1正则表达式检查 属性6非空检查 长度检查 TEST1表非空检查								
主题： 所有 客户 产品 渠道 交易 合作方 协议 事件								
系统： 所有 数据集市 数据仓库 核心系统								
重要程度： 所有 高 中 低								
							▶ 运行 ⬇ 规则导出	
规则ID	字段名	关联表	规则描述	规则模板	规则大类	负责人		
9	MAT_AT3_VALUE	TBLMDM_Z050_MASTER_PC	属性3取值范围检查	自定义规则检查	完整性	admin0	修改	删除
8	MAT_AT5_VALUE	TBLMDM_Z050_MASTER_PC	属性5符合下拉列表检查	自定义规则检查	完整性	admin0	修改	删除
6	MAT_AT4_VALUE	TBLMDM_Z050_MASTER_PC	属性4正则表达式检查	自定义规则检查	完整性	admin0	修改	删除
4	MAT_AT1_VALUE	TBLMDM_Z050_MASTER_PC	属性1正则表达式检查	自定义规则检查	完整性	admin0	修改	删除
3	MAT_AT6_VALUE	TBLMDM_Z050_MASTER_PC	属性6非空检查	空值检查方法	完整性	admin0	修改	删除

数据质量：对检核结果进行汇总分析

- 根据数据质量检核结果，从多个维度进行统计分析，发现企业系统、部门、加工链路中的薄弱环节。



数据标准：对企业各类标准进行统一展示和应用

- 企业基础标准、指标标准、公共代码等各类标准，需要通过统一的平台进行发布，以保证全企业都可以查询和使用。

业务数据标准查询			
主题域: <input type="text" value="-请选择-"/>	类型: <input type="text" value="-请选择-"/>	名称: <input type="text"/>	
业务定义: <input type="text"/>	编码(号): <input type="text"/>	<input type="button" value="搜索"/>	
第1页,共132页 首页 下一页 末页 每页10条,共有1311条记录 <input type="text"/> <input type="button" value="Go"/>			
名称	类型	定义	路 径
三年平均销售额	业务数据	指客户三年以来年平均销售额。	/客户个人客户财务信息/单位客户扩展信息/三年平均销售额
三年平均营业额	业务数据	指客户三年以来年平均营业额。	/客户个人客户财务信息/单位客户扩展信息/三年平均营业额
低风险客户标志	业务数据	客户所发生的业务都为低风险业务类型的,则该客户为低风险客户。	/客户个人客户风险信息/单位客户风险信息/低风险客户标志
新资本协议客户类型	业务数据	指新资本协议中规定的客户的类型,覆盖保证人的客户类型、有价证券的发行人客户类型、交...	/客户个人客户风险信息/单位客户风险信息/新资本协议客户...
企业信用评级	业务数据	指招行对企业信用的评级。	/客户个人客户风险信息/单位客户风险信息/企业信用评级
首次建立授信关系日期	业务数据	指客户在招行第一次发生授信业务的日期。	/客户个人客户往来信息/单位客户往来信息/首次建立授信关...
首次建立一般风险授信...	业务数据	指客户在我行首次发生一般风险业务的日期。	/客户个人客户往来信息/单位客户往来信息/首次建立一般风...

数据标准：实现对数据标准的规范管理

- 企业数据标准相关的管理规范，如标准变更的流程等，通过系统功能保障其实现。

The screenshot displays a web-based interface for managing data standards. At the top, a horizontal progress bar indicates the workflow stages: 创建申请 (Create Application), 保存提交申请 (Save and Submit Application), 选择主办人 (Select Organizer), 选择评审人 (Select Reviewer), 评审 (Review), 发布 (Publish), and 结束 (End). The '评审' (Review) stage is currently active, indicated by a blue dot and a blue circle.

Below the progress bar, the task name is '申请更新指标标准' (Apply for updating indicator standards). To the right of the task name, there are several status buttons: '任务信息' (Task Information), '全部通过并完成审批' (All passed and completed approval), '审批完成' (Approval completed), '修改评审人' (Modify reviewer), '任务记录' (Task record), '打印' (Print), and '返回我的任务' (Return to my tasks).

On the left side, there is a sidebar menu titled '数据标准' (Data Standard). It contains two main categories: '新增' (New) and '修改' (Modify). Under '新增', there are four items: '术语和缩略语(0)' (Terms and Abbreviations (0)), '业务数据标准(0)' (Business Data Standards (0)), '指标标准(0)' (Indicator Standards (0)), and '产品标准(0)' (Product Standards (0)). Under '修改', there are four items: '术语和缩略语(0)' (Terms and Abbreviations (0)), '业务数据标准(0)' (Business Data Standards (0)), '指标标准(1)' (Indicator Standards (1)), and '产品标准(0)' (Product Standards (0)).

The main content area shows the '任务信息' (Task Information) tab. It contains the following details:

- 任务名称: 申请更新指标标准
- 任务附件: [datastandard_20150408.xls](#)
- 任务描述:



CONTENTS

目录

01

大数据时代数据的特点

02

大数据时代的数据治理

数据治理咨询

数据治理产品



数据治理实施

03

典型案例

覆盖全企业的数据治理实施工作

- 数据治理是一项“规划+产品功能+实施”缺一不可长期系统性工作，数据治理要取得成效，最终还是要落地到实施层面。
- 数据治理的落地实施往往不是一个系统、一个部门的事，需要企业多个或者全部部门共同参与。
- 数据治理的实施不仅包括功能层面的，还包括内容层面的，比如数据标准的持续更新，数据质量检核规则的整理和实施等。



- 数据标准是全企业需要共同遵循的规范，其建立过程中需要相关部门共同进行讨论和审批；而标准一旦发布，也需要相关部门严格遵守，以此规范新建系统以及有条件的改造旧系统。

序号	标准编号	主题	数据项一级分类	数据项二级分类	数据项三级分类	中文名称	关键字	英文名称	业务定义	基础标准类型
97	BS170097	客户	单位客户	基本信息		证件类型 (对公)	证件类型 (对公)	Type of Corporation	用于单位身份识别，依法设立	代码型
98	BS170098	客户	单位客户	基本信息		组织证件号码	组织证件号码，证件	Number of	用于单位身份识别，依法签发	编码型
99	BS170099	客户	单位客户							

核心系统中字段与标准信息项匹配情况表

说明：存在差异的部分采用黄色背景色标识

序号	信息项中文名称		信息项英文缩写名		数据类型		数据长度	
	系统	标准	系统	标准	系统	标准	系统	标准
1	empId	员工工号	empId	EMP_ID	VARCHAR2	VARCHAR2	64	8
2	姓名	员工姓名	empName	EMP_NAME	VARCHAR2	VARCHAR2	64	50
3	显示顺序	显示顺序	displayOrder	DSPLY_ORDE	integer	CHAR	0	4
4	机构ID	机构号	orgId	ORG_ID	VARCHAR2	VARCHAR2	64	30
5	机构名称	机构名称	orgName	ORG_NAME	VARCHAR2	VARCHAR2	64	100
6	部门Id	部门号	deptId	DEPT_ID	VARCHAR2	VARCHAR2	64	30
7	部门名称	部门名称	deptName	DEPT_NAME	VARCHAR2	VARCHAR2	64	100

数据治理实施——质量问题发现及解决

- 数据质量与企业众多部门和系统密切相关，数据质量的检核规则也需要各部门/系统集成思广益，对经常出现问题或者重要的环节进行质量监控；对于发现的数据质量问题，各部门/系统需要积极进行整改，才能保证数据质量不断提升。

项目部门	数据源名称	表名	字段名	主题	规则描述	源系统					
信管部	MySql				必输字段空比率(%)						
信管部	MySql		2018年第三季度数据质量检查问题清单								
信管部	MySql		完整性	准确性(有效性)	唯一性	一致性	合规性	三季度合计	二季度合计	备注	
信管部	MySql		ECIF	11	67	4	26	0	108	85	下降3%
信管部	MySql		对公信贷系统	0	19	3	4	0	26	66	提升13%
信管部	MySql		个贷系统	2	17	8	2	8	37	45	提升2%
信管部	MySql		核心系统	3	3	12	4	2	24	27	提升0.5%
信管部	MySql		人力资源系统	2	3	1	4	9	19	20	提升0.5%
信管部	MySql		资金交易系统	5	11	2	3	15	36	41	提升0.5%



CONTENTS

目录

01

大数据时代数据的特点

02

大数据时代的数据治理

03

典型案例



案例1：某大型国有银行数据申请服务

- XX行数据申请服务管理是将原来各分行通过邮件进行数据申请的过程在数据标准平台上规范化；
- 数据标准平台对全部系统数据字典进行定期、及时的更新，保证其反映各系统现状；分行和总行其他部门用户可以通过数据字典选择需要的数据并提出数据申请。

数据标准管理系统

你好: admin 退出 帮助

数据文件申请

保存草稿 提交 关闭

基本信息

申请分行: --

数据接收系统名称:

数据接收系统代码:

数据申请详细

1 请选择申请对象

表名:

增量标识: --

* 数据文件名:

* 数据接收目录:

* 下载起始日期:

新增数据申请

附件上传

分行业务部门相关附件

添加附件

分行技术部门相关附件

添加附件

选择数据字典

数据字典

国内

海外

AAOP

AAS

AASCBO

ASTARI

ASTARO

ASTBGL

ASTCAD

ASTCCY

ASTCCYF

ASTFFX

ASTINP

ASTINR

ASTINV

ASTIRT

ASTOBH

ASTOBP

ASTORG

ASTPAC

ASTPEN

ASTPTX

取消 确定

申请日期: 2016-12-21

数据接收用户:

源系统: --

数据源子分省: --

* 数据使用系统:

* 数据下载时间要求:

案例1：某大型国有银行数据申请服务

- 数据申请需要经过分行技术/业务部门、分行数据主管部门、总行主管部门等多个层级的审批，规范了企业内部数据分发的过程；
- 定期对经常被申请的数据、分行申请数据量、分行申请数据是否重复等进行分析，以便更好的提供相关服务。

The screenshot displays the '数据标准管理系统' (Data Standard Management System) interface. The top navigation bar includes a search bar, a welcome message '欢迎：本行审批', and buttons for '退出' (Logout) and '帮助' (Help). The main content area is titled '数据文件申请详情' (Data File Application Details) and includes a '关闭' (Close) button. The application details are as follows:

申请分行:	联系人/部门/电话:	申请日期:
河北	本行审批/a/011-7867347	2017-08-08

数据接收系统名称:	数据接收系统IP:	数据接收用户:
qq	192.169.1.168	qq

数据接收密码:
qwe123

1	表名:ASTARI	源系统:不规范个人账户处理系统	表中文名: ASTARI	数据提出部门:qq	操作
					[+展开]

分行数据应用特色业务系统需求申请书、分行申请部门审批意见（均需提供）

文件名称
20170807103022...

总行数据业务主管部门对数据文件使用的批复意见（申请不分省数据需要提供）

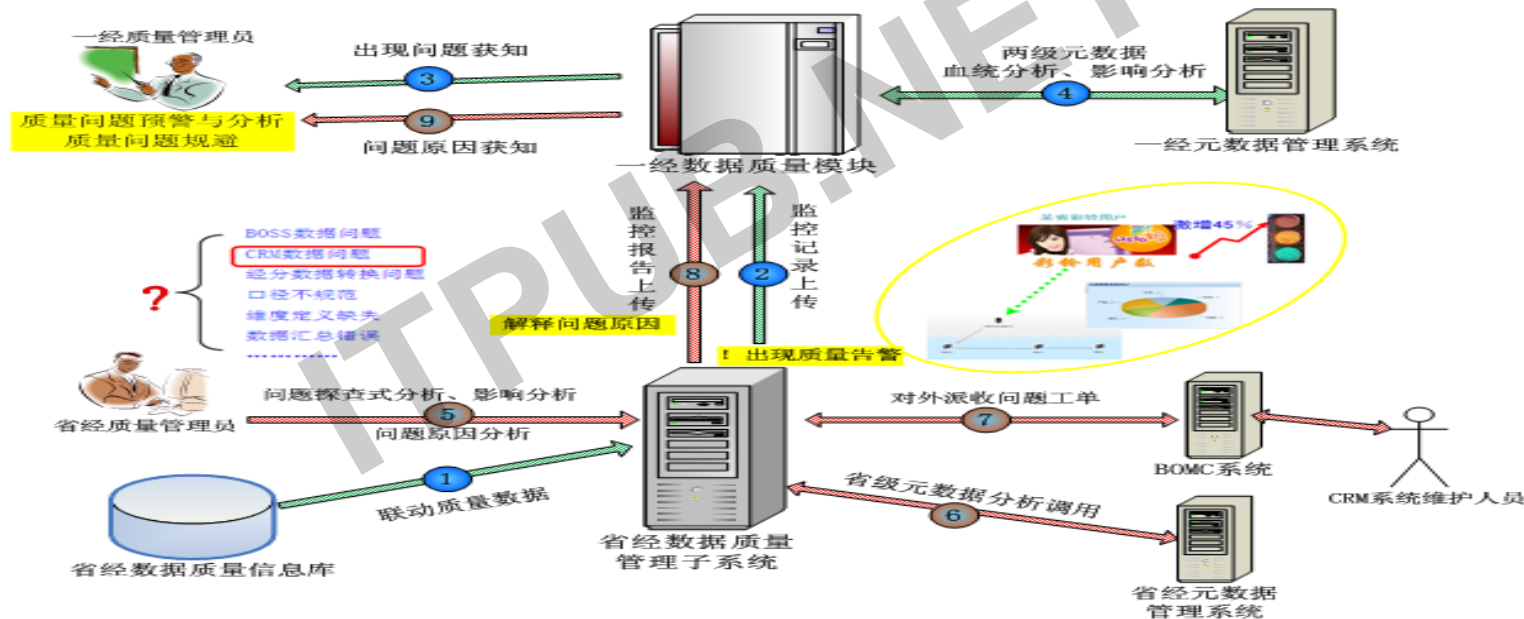
文件名称
20170807103022...

分行数据申请服务主管部门审核意见

[添加附件]

案例2：某电信运营商数据质量两级联动

- 某运营商集团要求各省各公司定期上报关键KPI指标数据并进行检核，发现的数据质量问题也会责成各省分析原因并进行解决。



案例2：某电信运营商数据质量两级联动

- 在各省数据质量系统上可以设置和查询上报集团的时间、上报内容、是否成功上报等信息。

QOP挂裁功能 | 刷新

- 联动信息
 - 监控报告
 - 监控记录
 - 上传下发文件
- 联动调度
 - 调度配置
- 编码维护
 - 告警分层管理
 - 告警级别管理
 - 告警原因类型管理
 - 告警状态类型管理
 - 监控规则类型管理
 - 接口范围管理
 - 联动及时性监控规则管理
 - 移动省公司机构代码管理
 - 指标范围管理
- 系统管理
 - 操作记录查询

当前位置: 联动信息定制 > 联动信息 > 监控记录

查询条件

开始日期: 2012-03-01 结束日期: 2012-03-20

一经接口单元名称: 全部 一经指标名称: 全部

告警状态类型名称: 全部 监控规则类型名称: 全部

告警分层类型名称: 全部 告警分级类型名称: 全部

查询

推送选中信息

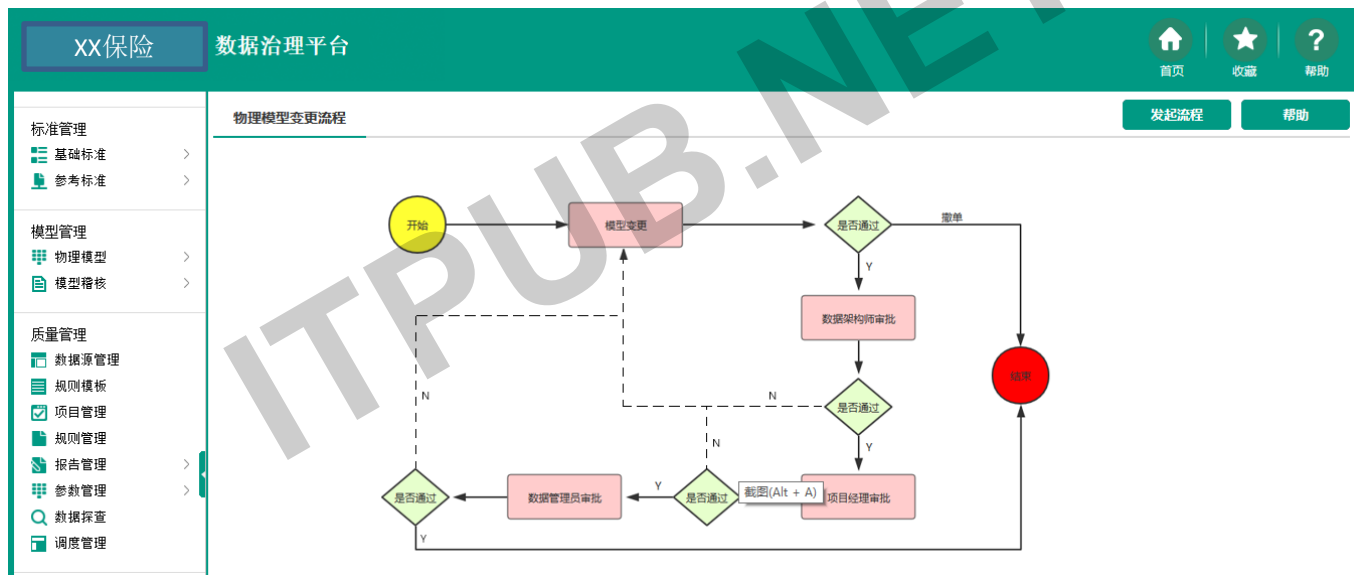
导出为Excel文件 导出为TXT文件

全选	推送状态	一经质量监控对象编码	一经质量监控对象编码名称	告警时间	数据日期	省经监控对象元数据ID	省经监控对象监控规则类型编码	省经监控对象监控规则类型编码名称	一经
<input checked="" type="checkbox"/>	已推送	2207331212	联通移动客户总数	2012-03-02	2012-03-15	META=p2	211	同比波动规则	
<input type="checkbox"/>	未推送	2220108212	使用TD网络的数据卡客户数	2012-03-03	2012-03-15	META=p2	100	及时性规则	
<input type="checkbox"/>	未推送	0204700110	用户使用终端通话情况	2012-03-04	2012-03-15	META=p2	110	及时性规则	
<input type="checkbox"/>	未推送	2203813212	当日短信类业务收入	2012-03-07	2012-03-14	META=p2	210	波动规则	
<input type="checkbox"/>	未推送	2220109212	使用TD网络的手机客户数	2012-03-08	2012-03-14	META=p2	240	自定义规则	
<input type="checkbox"/>	未推送	2203811212	当日增值业务收入	2012-03-12	2012-03-13	META=p2	200	准确性规则	
<input type="checkbox"/>	未推送	2207326212	联通移动新增客户数	2012-03-06	2012-03-13	META=p2	220	平衡规则	
<input type="checkbox"/>	未推送	2203805212	当月累计通信客户数	2012-03-14	2012-03-12	META=p2	110	及时性规则	

第 1 页/共3页 每页8条/共18条

案例3：某保险公司数据模型标准化

- 将数据标准应用于企业现有建模工具，要求新建模型时必须应用已有数据标准，并对其进行检查，检查通过后，才在系统中创建相关表。



案例3：某保险公司数据模型标准化

- 在系统中不仅对模型变更进行审批，还可以查看所有已有模型以及变更记录。

XX保险 数据治理平台

模型变更

标准管理

基础标准

参考标准

模型管理

物理模型

模型稽核

质量管理

数据源管理

规则模板

项目管理

规则管理

报告管理

参数管理

数据探查

调度管理

模型变更历史记录

导出 返回

模型CODE: 123 模型名称: 12 系统名称: 核心业务系统LIS6.0
变更人: admin1 变更时间: 2019-03-19 15:38:08 对应Schema: LISPRDD
模型变更说明:

变更Table列表

序号	表名	表中文名	表备注	表主键	变更类型
1	LASPayPerson	渠道应收个人交费表	个险使用, 渠道使用, 可废弃	PolNo, PayCount, PayAimClass, DutyCode, PayPlanCode, PayType	新建
2	LDUWDifferenceB	业务员差异化核保级别表	个险使用	DiffType, SerialNo	新建
3	LACommisionDetailB	保单发佣分配份表	个险使用	EdorNo, GrpContNo, AgentCode	新建
4	LACommisionDetail	保单发佣分配表	个险使用	GrpContNo, AgentCode	新建
5	LAAgentGreenChnl	绿色通道代理人信息表	绿色通道代理人信息表	AgentCode, StateType	新建

石竹公司简介

➤ 石竹软件：数据治理领域的专业软件开发商及服务提供商。

- 全称：石竹计算机软件有限公司
- 创建于1999年
- 办公地点：北京、西安、广州、上海

➤ 始终专注，业界领先

- 15+年的数据治理垂直领域的专注发展；
- 元数据产品在中国市场占有率第一；
- 国际CWM规范组织参与者；
- 国际数据质量协会IAIDQ大中华区合作伙伴；
- 赛迪顾问“行业用户首选元数据/数据质量管理产品与解决方案”。



微信:

联系方式:

姓名: 薛勇

手机: 13911870318

邮箱: 13911870318@139.com





THANKS

ITPUB3.NET