

大数据治理解决方案

中国西部最大的CIO交流平台

网址: www.sccio.cn

公众号: 成都企业信息化促进会

提供IT技术方案与实战案例PPT、
供需对接、人才、技术培训等服务

四川CIO俱乐部 www.sccio.cn
成都企业信息化促进会

目录

1

大数据管控概述

2

元大数据管理

3

大数据标准管理

4

大数据质量管理



大数据管控系统实施目标

大数据管控系统实施是为解决企业所面临的大数据标准问题、大数据质量问题、元大数据管理问题。

大数据标准规范落地

- 推动大数据标准在全企业的执行落地，规范化管理构成大数据平台的业务和技术基础设施，包括大数据管控制度与流程规范文档、信息项定义等

大数据质量问题定位分析

- 全方位管理大数据平台的大数据质量，提升大数据可访问性、可用性、正确性、一致性等，实现可定义的大数据质量检核和维度分析，以及问题跟踪。

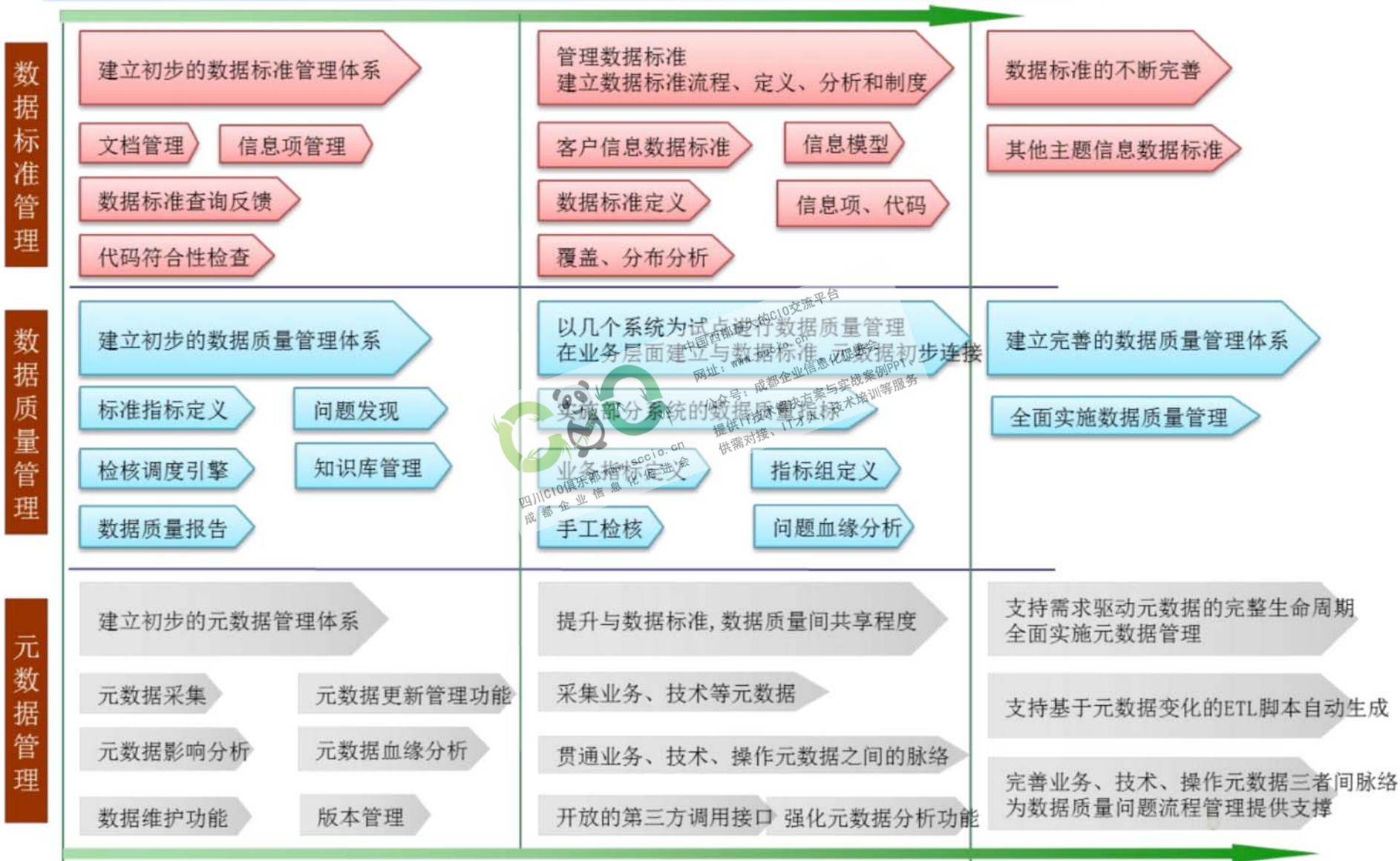
大数据关系脉络化

- 提升统一有序的业务系统和MIS系统的管理大数据能力，实现对大数据间流转、依赖关系的影响和血缘分析，增加有效工作时间用于分析大数据，减少用于在复杂企业大数据环境中搜索大数据的时间

大数据管控体系建设原则



大数据管控系统实施步骤



目录

1

大数据管控概述

2

元大数据管理

3

大数据标准管理

4

大数据质量管理



元大数据定义

Synonyms

- 元数据的英文名称是“Metadata”，它是“关于数据的数据”，通俗的说法是描述性数据。

我们可以根据需要，从不同的角度去描述一个数据的元数据，例如：

“3. 14159265358979...”

对这样一个数据，我们可以从下面一些角度去描述它的元数据

名称：圆周率

别名： π

定义：圆周长与直径的比值

数据类型：数学常量

相关人物：祖冲之、韦达……

Definition

Age

Size

Owner

Location

Version

Security Level

Data Type

Relationships to
Other Data

元大数据分类

元数据描述了**数据类系统的结构和建立方法**。

- 元数据为访问数据类系统提供了一个**信息目录** (*InformationDirectory*) ,
 - » 它全面描述了系统中都有什么数据、这些数据怎么得到的、如何访问这些数据。
 - » 它是数据系统运行和维护的中心。系统服务器利用他来存贮和更新数据, 用户通过他来了解和访问数据。
- 按用途不同分为**技术元数据**、**业务元数据** 和 **操作元数据**
 - 技术元数据: 设计、管理人员开发和管理数据系统的数据;
 - 业务元数据: 从商业业务的角度描述系统数据的数据;
 - 操作元数据:

元大数据范围

- 技术元数据包括：

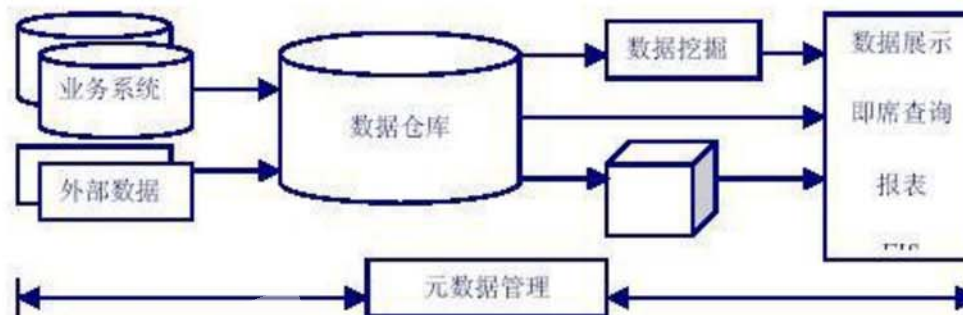
- » 数据源信息；
- » 源数据到目的数据的映射；
- » 数据转换的描述；
- » 数据仓库内对象和数据结构的定义；
- » 数据清理和数据更新时用的规则；
- » 用户权限，数据备份与导入的历史，信息发布历史等。

- 业务元数据包括：

- » 业务主题的描述；
- » 包含的数据、查询、报表；
- » 指标。

- 操作元数据包括：

- » ETL加载过程的状态，结果；
- » 包含的警告，读写信息；
- » 作业信息。



中国西部最大的CIO交流平台
网址: www.sccio.cn

本公众号: 成都企业信息化促进会
提供IT技术解决方案与实战案例PPT、
供需对接、IT培训等服务

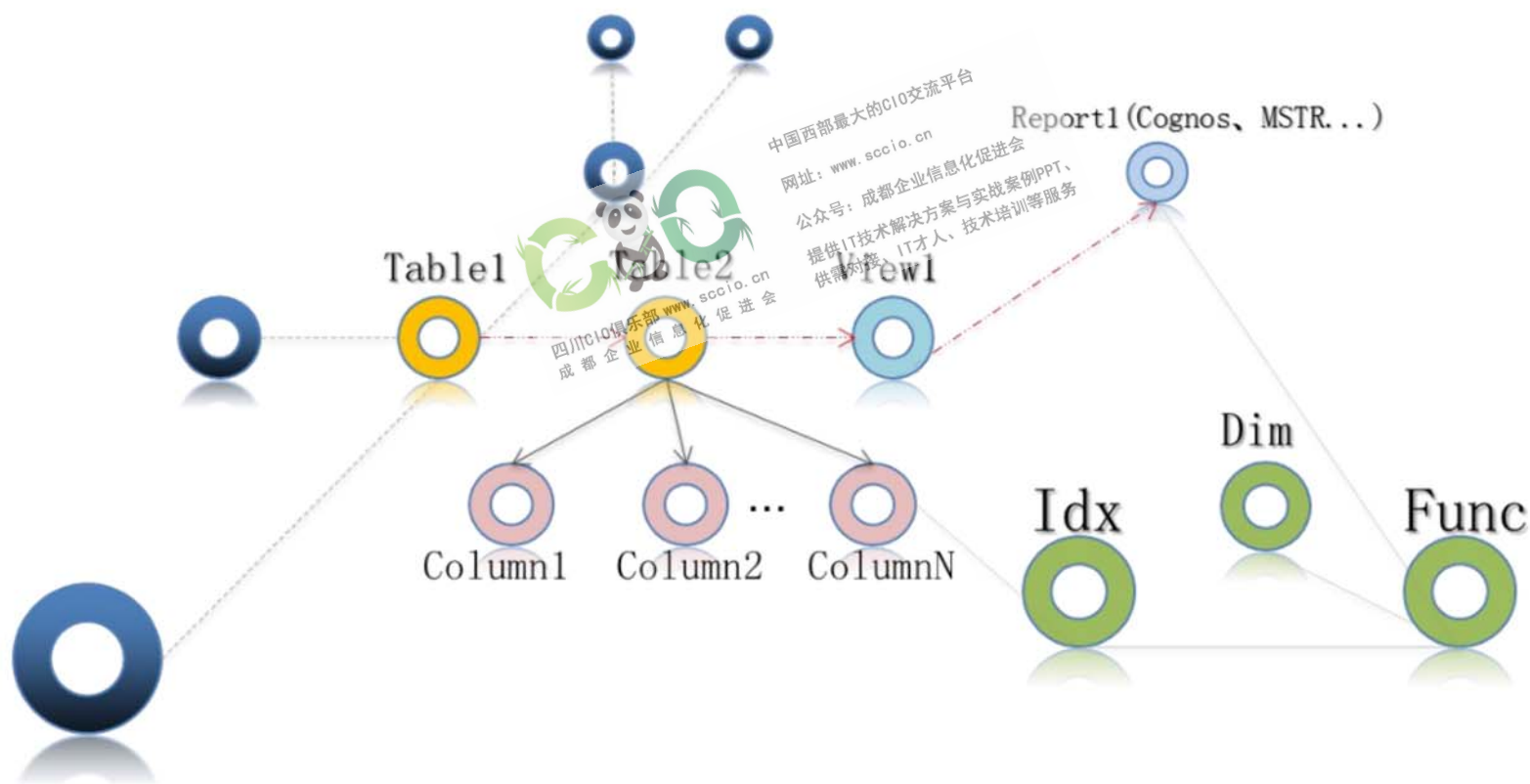
四川CIO俱乐部 www.sccio.cn
成都企业信息化促进会

元大数据关联性

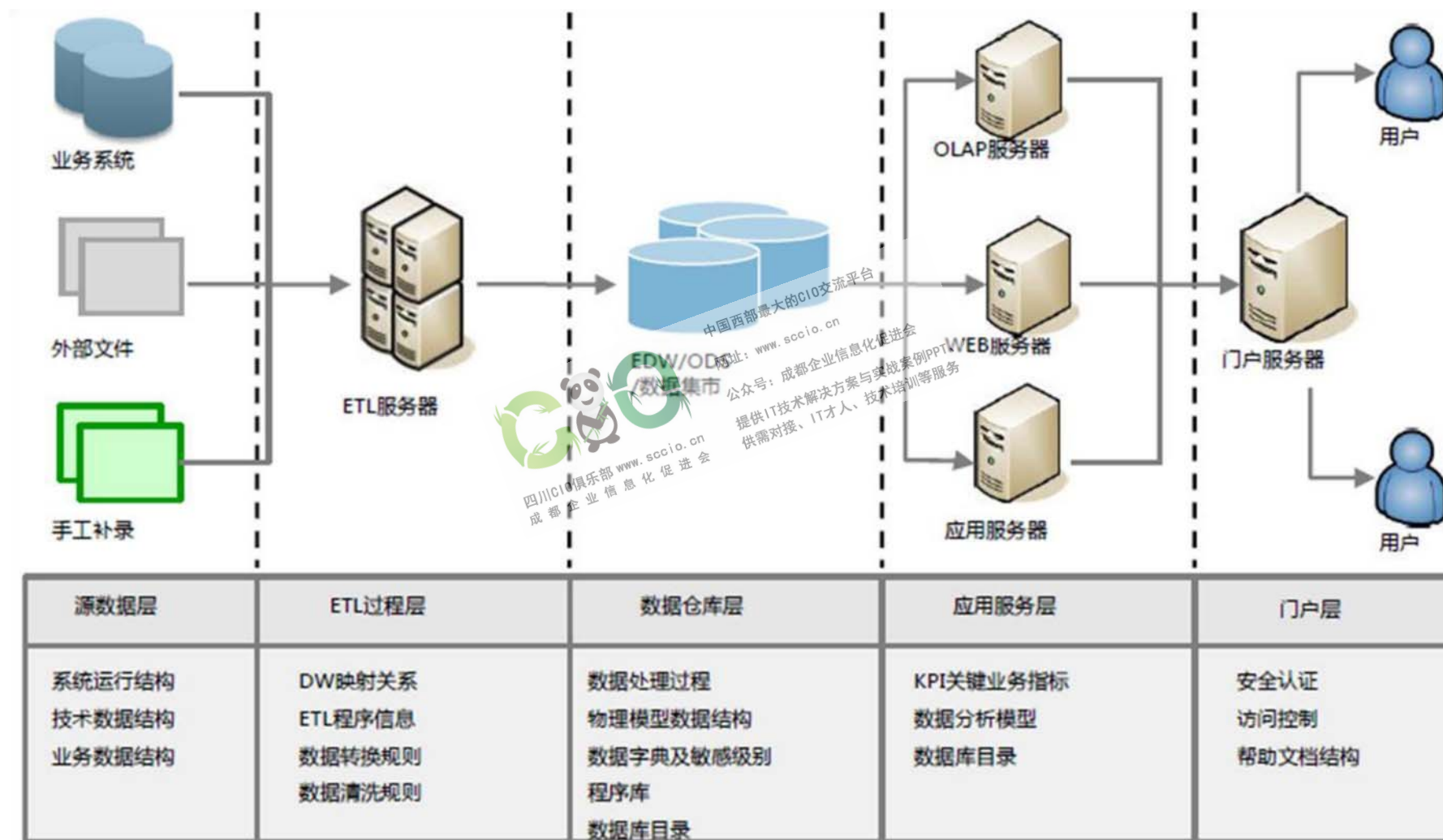
元数据管理除了能展示元数据的实体信息外，还能查看与其有关联的其它元数据的信息；

✦ 不同类型之间的元数据，相同类型但不同的元数据之间，都可以存在这种关系；如库与表、表与字段之间的包含关系，表到表、表到报表、字段到字段之间的加工关系；

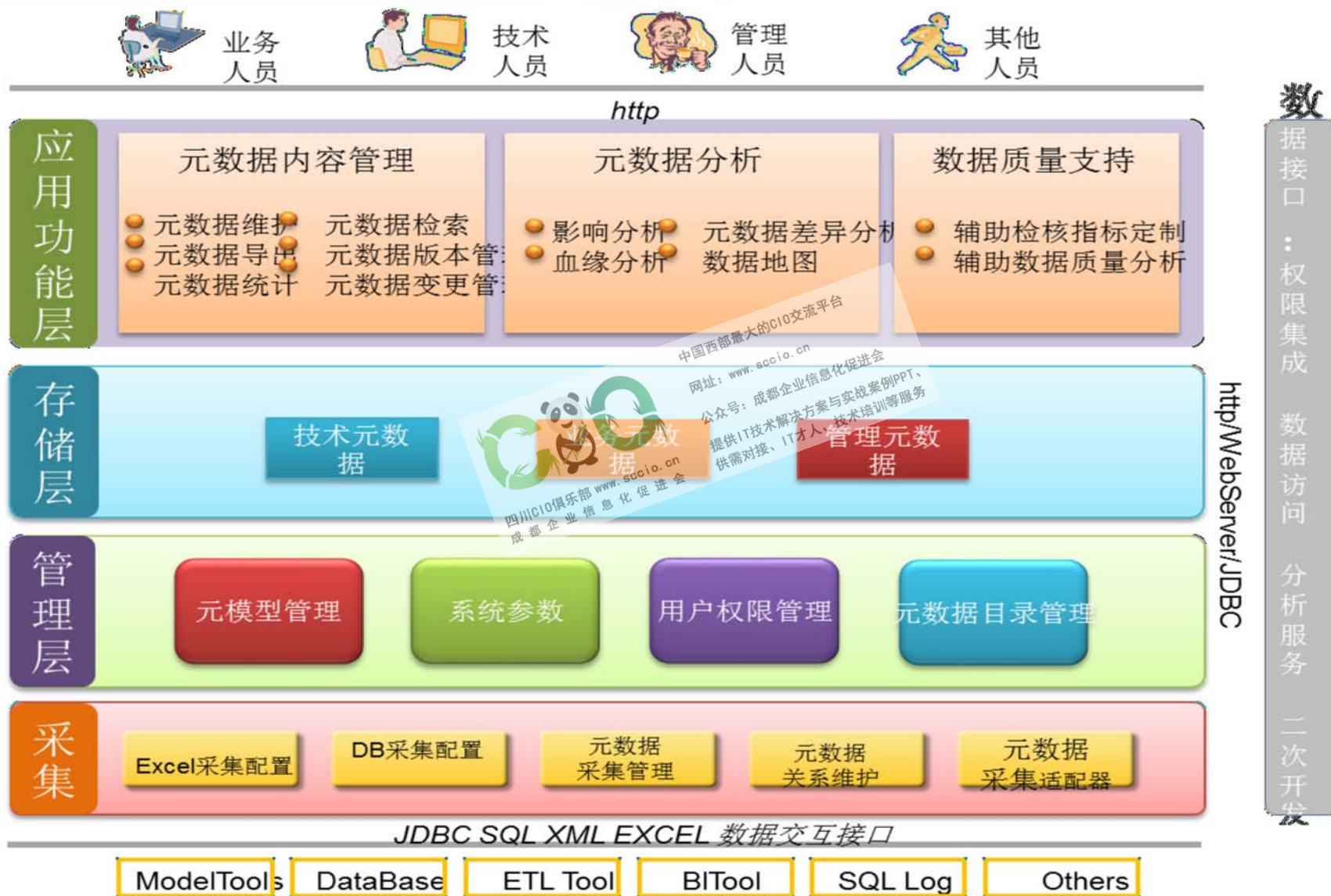
✦ 借助可描述的各种关系，将大量貌似纷繁芜杂、毫无规律的元数据组织成一个有序的网络拓扑图；



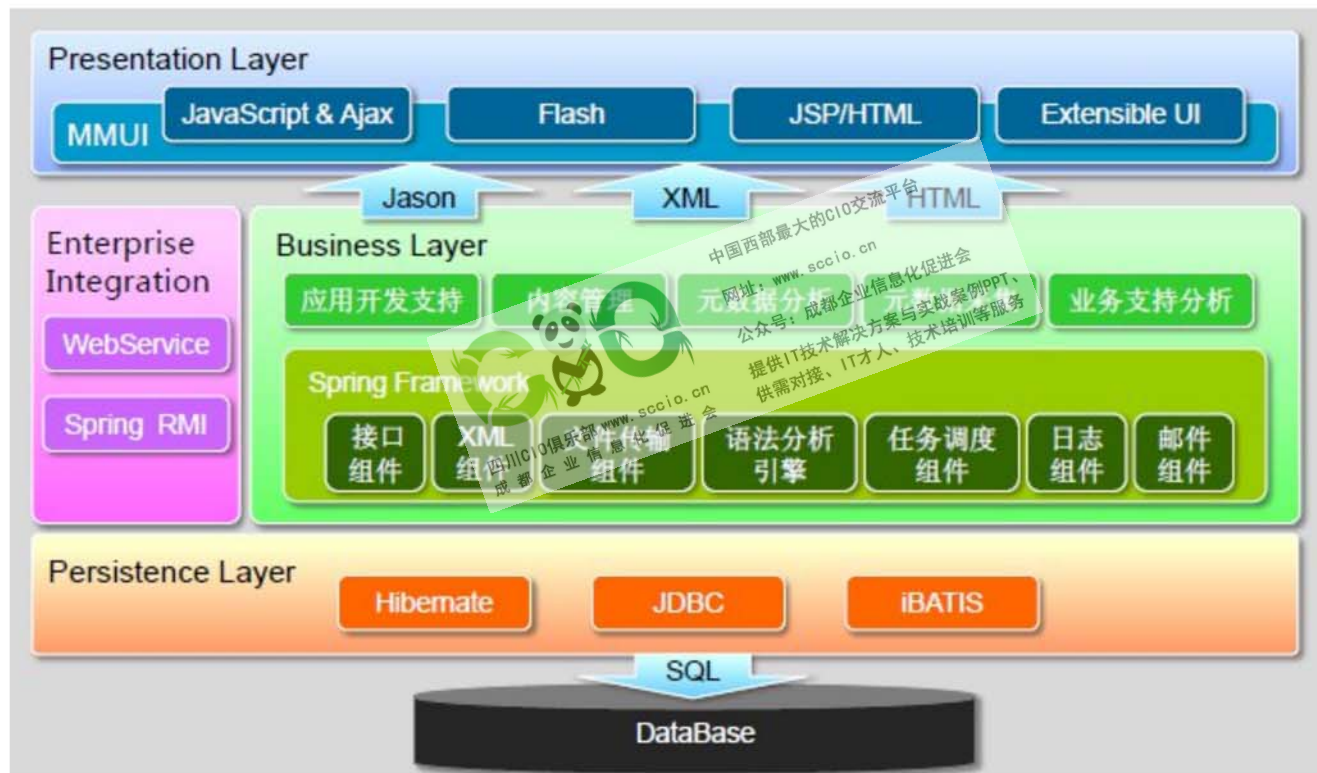
元大数据系统应用模型



元大数据管理系统功能构成图



技术架构图



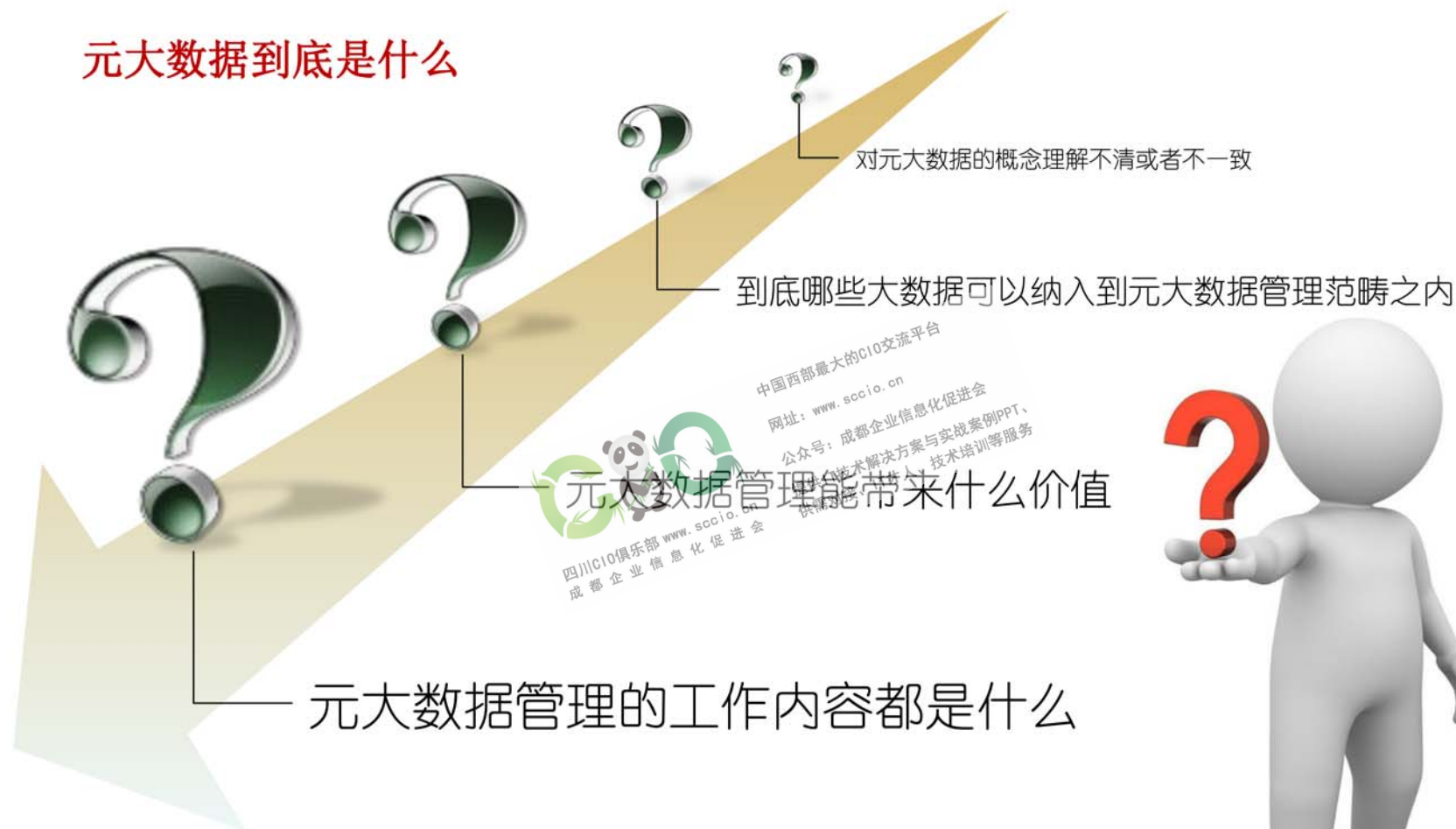
元大数据管理-现状分析

企业大数据资产管理的困惑



元大数据管理-元大数据的意义

元大数据到底是什么



中国西部最大的CIO交流平台
网址: www.sccio.cn
公众号: 成都企业信息化促进会
提供: 技术解决方案与实战案例PPT、
技术培训等服务

四川CIO俱乐部 www.sccio.cn
成都企业信息化促进会

元大数据系统-解决方案

解决之道



需要借助元大数据管理来解决理想和现实之间的差异问题！

- ✦ 元大数据能够辅助管理企业的各类大数据资产；
- ✦ 元大数据可以如实向用户反映企业的大数据资产信息；
- ✦ 元大数据管理的工作内容包括人员组织搭建、管理流程的梳理、基础技术平台的实施；
- ✦ 元大数据的解决方案回答了如何让元大数据管理产生价值的问题。

元大数据系统-元大数据管理定位

⊕ 定位：

元大数据管理是对大数据平台大数据信息的梳理、组织和再现，帮助用户更好的理解现有系统的建设现状，支撑用户对大数据平台的管理工作作出更合理的决策，但不能过分夸大元大数据管理的工作范围，它并不能替代现有大数据平台开发和管理工具的角色，也不能彻底改变现有大数据平台的

中国成都最大的CIO交流平台
网址: www.sccio.cn
公众号: 成都企业信息化促进会
提供: 技术解决方案与实战案例PPT、
技术培训等服务

四川CIO俱乐部 www.sccio.cn
成都企业信息化促进会



元大数据系统-价值

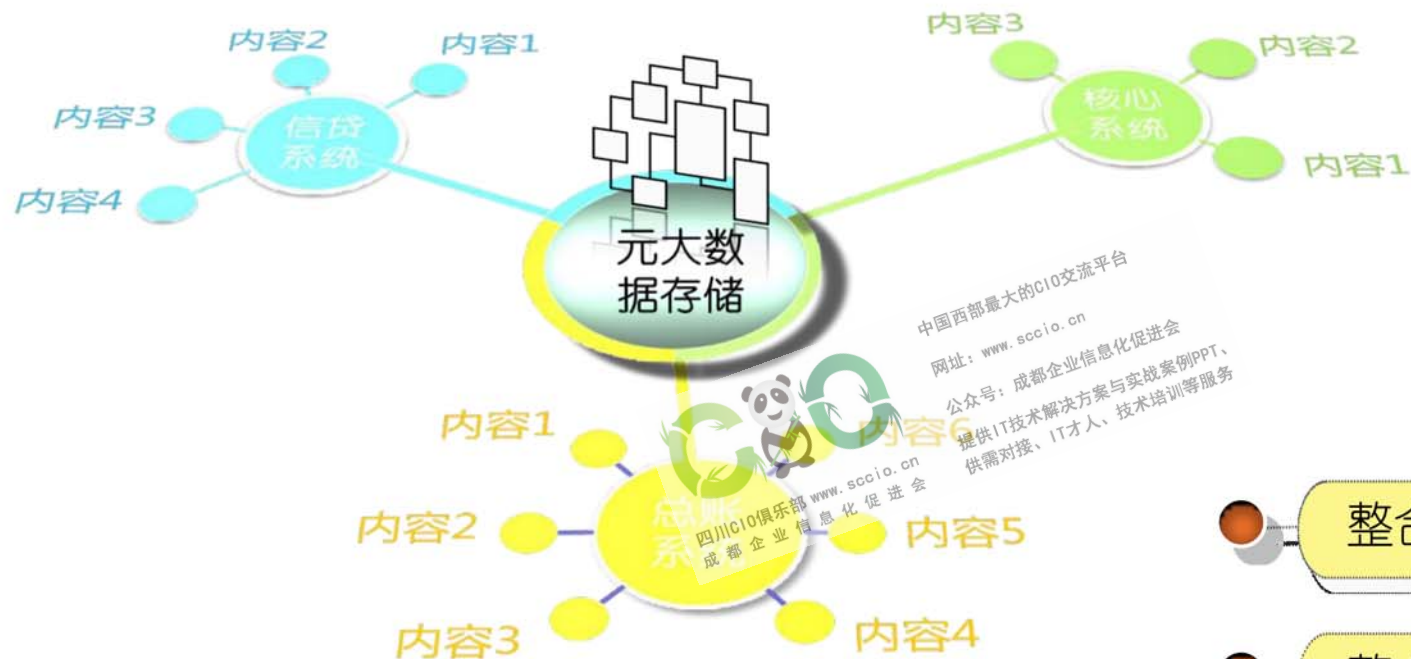
⊕ 元大数据管理的应用价值

元大数据管理的应用价值主要体现在：

- ❑ 对大数据再组织并形成全局性的视图；
- ❑ 帮助用户更好的理解各环节的大数据和系统的建设现状；
- ❑ 是保障企业大数据质量的基础；
- ❑ 支持企业信息化的知识传承；
- ❑ 提升大数据平台建设和管理水平。



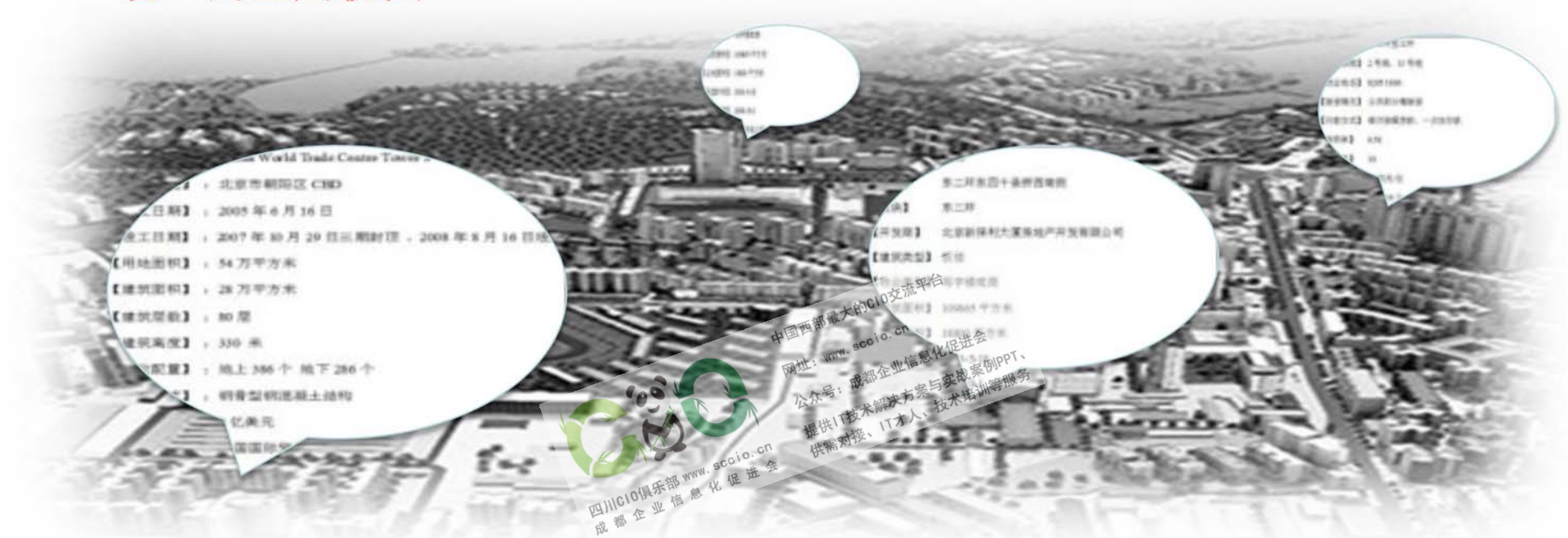
元大数据管理-整合



- 整合不同系统的元大数据
- 整合不同来源的元大数据
- 整合不同类型的元大数据
- 统一可维护的元大数据存储

元大数据管理-大数据地图

统一的全局视图



元大数据解决方案-元大数据应用信 息



- 在查看报表的同时辅以文字化的术语说明；
- 提供统一、清晰的业务定义和口径；
- 是业务人员理解大数据的好帮手。

元大数据解决方案-信息检索

搜索中心

用户搜索 关键字: 证券账户 搜索类别: 全部 勾选搜索 点击搜索

应用功能搜索结果

功能名称	功能描述	类型	路径
证券账户数和投资者数趋势	证券账户数和投资者数趋势	仪表分析	X2I
证券账户数和投资者数趋势_014	证券账户数和投资者数趋势	仪表分析	X2I
证券账户资料查询	查询指定账户的资料并提供账户卡打印功能	报表	公共页面>业务功能>投资者与账户
证券账户数和投资者数	证券账户数和投资者数	灵活分析	X2I
证券账户数和投资者数	证券账户数和投资者数	灵活分析	X2I>参数自动刷新

元数据搜索结果

功能名称	功能描述	类型
证券账户组成情况	证券账户组成情况	业务需求
信用证券账户投资者分类统计	信用证券账户投资者分类统计	业务需求
合伙企业无合伙人证券账户信息查询	合伙企业无合伙人证券账户信息查询	业务需求
配售对象报备证券账户情况	配售对象报备证券账户情况	业务需求
基金部新增机构投资者证券账户	基金部新增机构投资者证券账户	业务需求
信用证券账户开户汇总	信用证券账户开户汇总	业务需求

数据标准搜索结果

功能名称	描述	类型
证券账户红利历史	指证券登记结算机构为投资者设立的,用于准确记载投资者所持的证券种类、名称、数量及相应权益和变动情况的账册,是认定股东身份的重要凭证,具有证明股东身份的法律效力,同时也是投资者进行证券交易的先决条件。	业务对象
证券账户		业务对象

中国西部最大的CIO交流平台
网址: www.sccio.cn
公众号: 成都企业信息化促进会
提供IT技术解决方案与实战案例PPT、供需对接、招聘、技术培训等服务

- 通过快速搜索获取元大数据信息;
- 支持包括中英文名称、关键字在内的搜索。

元大数据解决方案-大数据字典



- 通过快速搜索获取元大数据信息;
- 支持包括中英文名称、关键字在内的搜索。

元大数据解决方案-版本管理



- 支持自动定版和手工定版；
- 自动定版：元大数据信息发生变更，会自动维护版本；
- 手工定版：根据客户的需要将一到多个元大数据进行定版并可对版本命名；
- 不论自动还是手工定的版本，都可支持日后查阅和分析。

元大数据解决方案-影响分析

- 数据缓冲层的表结构是否需调整？
- 基础数据区的表结构是否需调整？
- 汇总层的表和视图结构是否需调整？
- 数据集市层的表、视图等结构是否需调整？
- 相关加工脚本是否需调整？
-



- 应用报表是否有效？
- 应用报表是否需调整？
-

- 表结构变更；
- 字段类型变化；
- 表删除
-

? 上游结构发生变动影响下游加工过程对象？

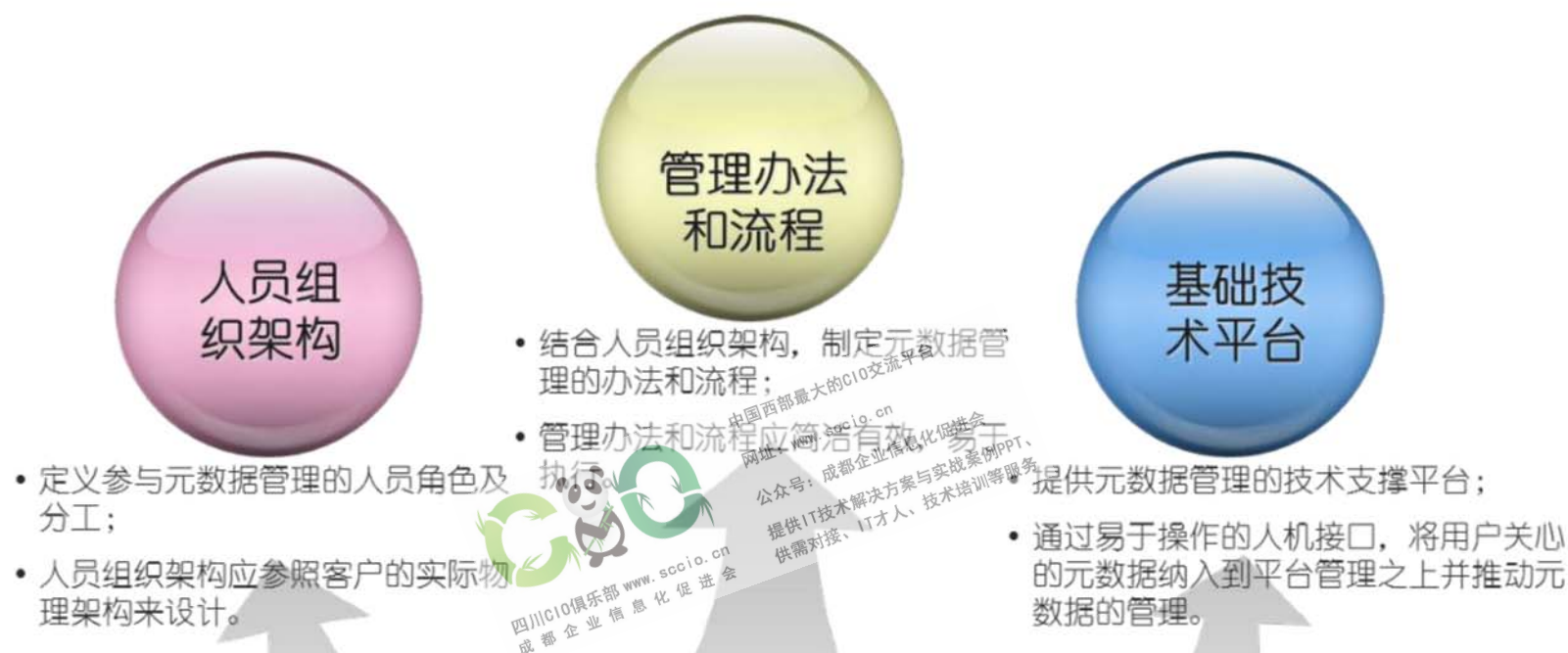
👍 通过影响分析获知变更对象列表并确认调整。

元大数据解决方案-大数据管控



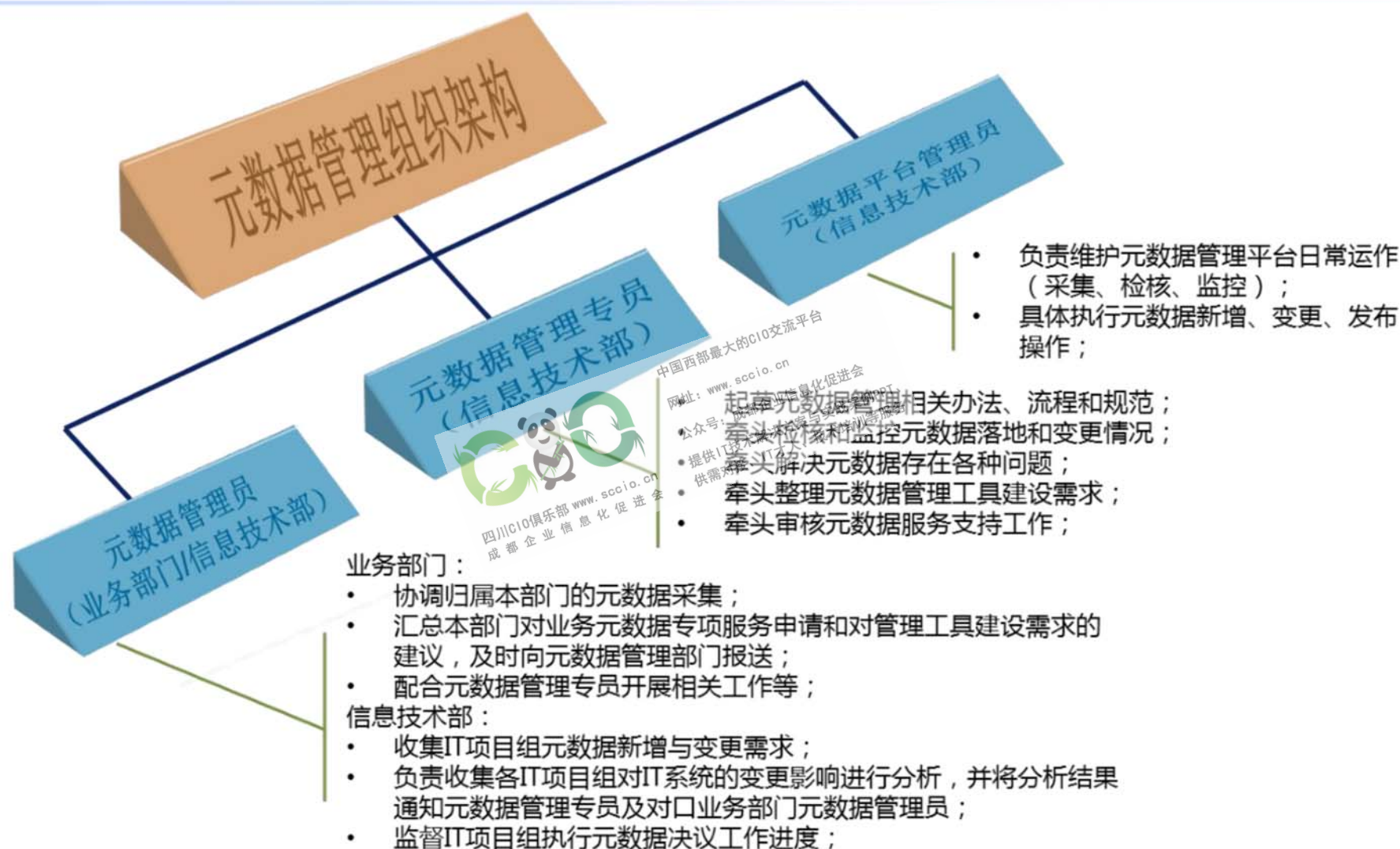
大数据管控 (Data Governance) 是指对一个企业大数据的可用性、实用性、完整性和安全性等的全面管理。一个有效的大数据管控包括一个管控主体、定义好的一组程序和一个执行这组程序的计划。

元大数据实施-总体规划

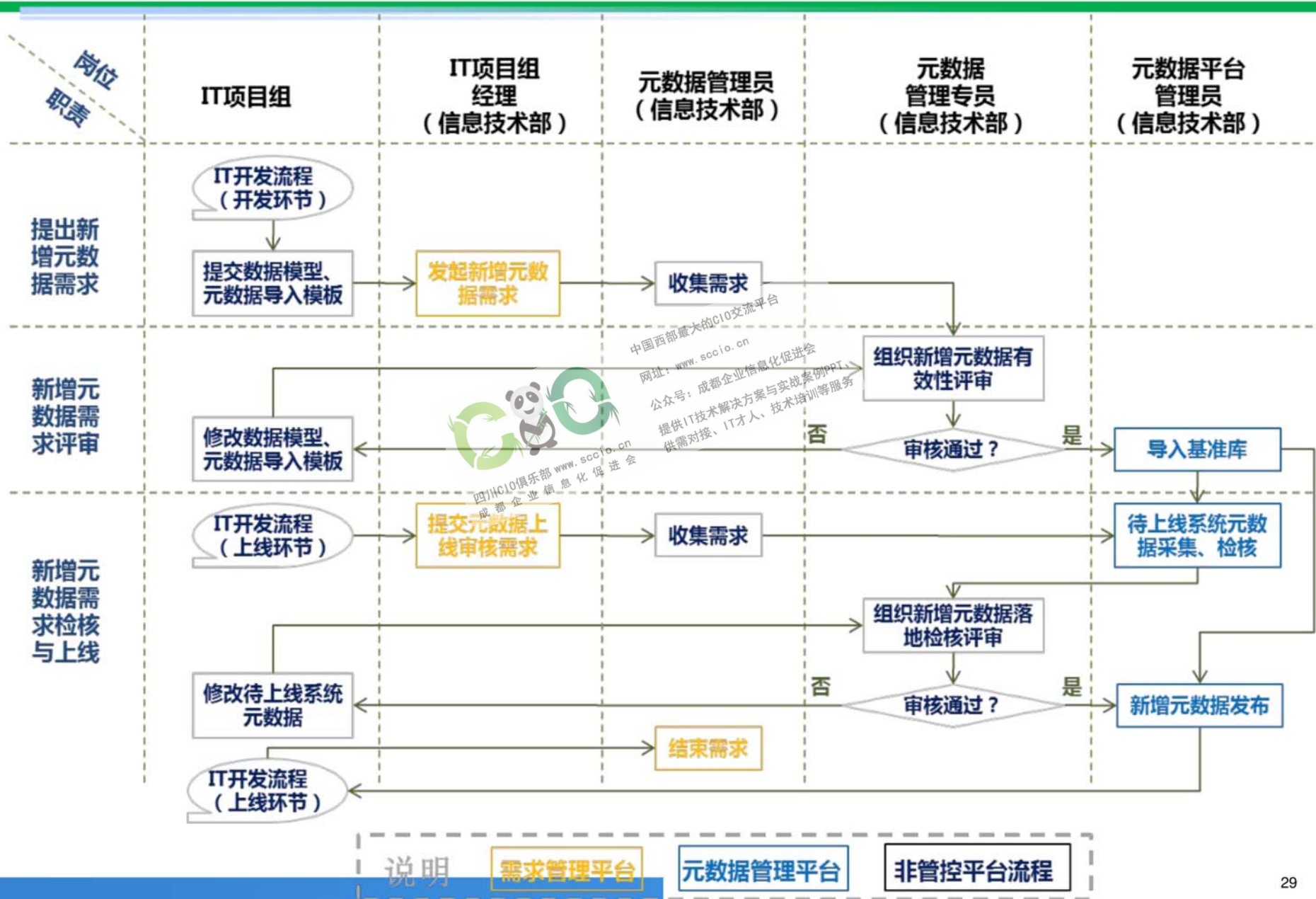


通过元数据管理办法与流程，信息技术部将统管全行的元数据（数据字典），负责组织、推动和协调元数据管理相关工作，包括提交元数据需求、组织元数据审批、元数据核检、元数据维护、元数据平台建设等；业务部门主要配合信息技术部开展上述相关工作。

元大数据实施-组织机构



元大数据实施-管理办法和流程



目录

1

大数据管控概述

2

元大数据管理

3

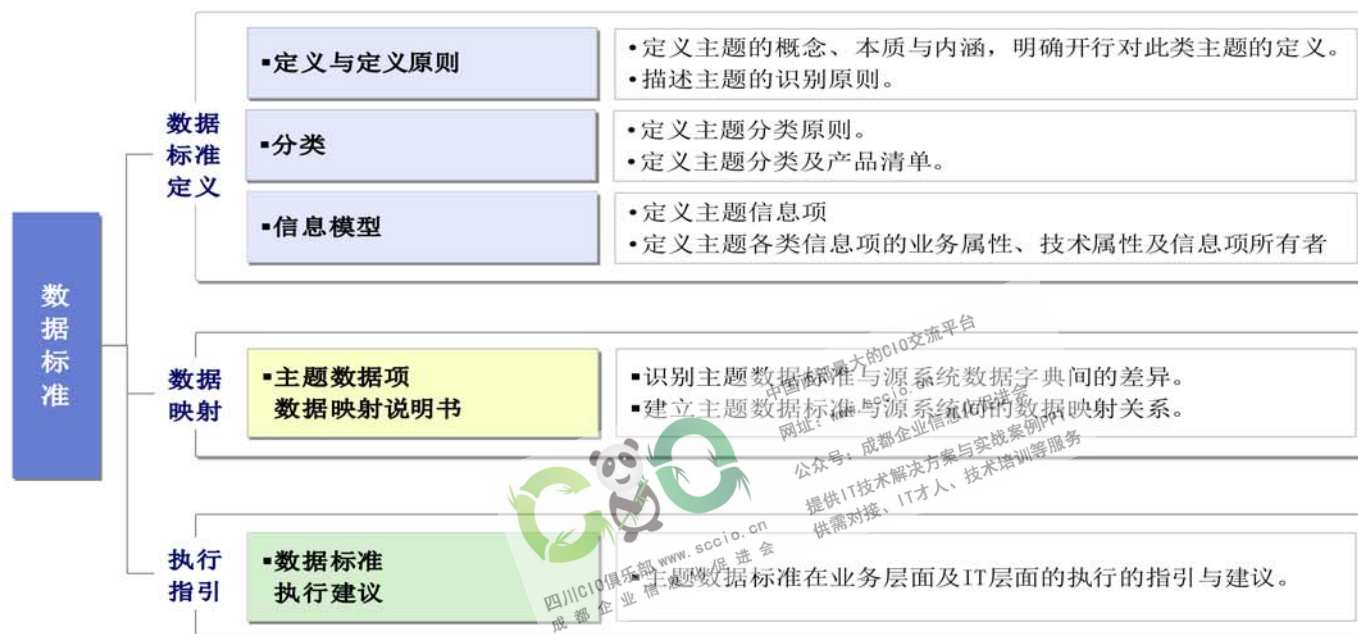
大数据标准管理

4

大数据质量管理



大数据标准管理



大数据标准管理应用场景



大数据标准工具逻辑架构图



大数据标准与元大数据关系

- 大数据标准是企业级的业务规范，用于指导各业务系统及大数据仓库的建设，而元大数据是系统级的描述手段，更多的反映系统建设情况；
- 大数据标准指导系统建设的成果可以通过元大数据来反映；
- 系统的建设反过来促进大数据标准的完善；

目录

1

大数据管控概述

2

元大数据管理

3

大数据标准管理

4

大数据质量管理



大数据质量问题产生原因



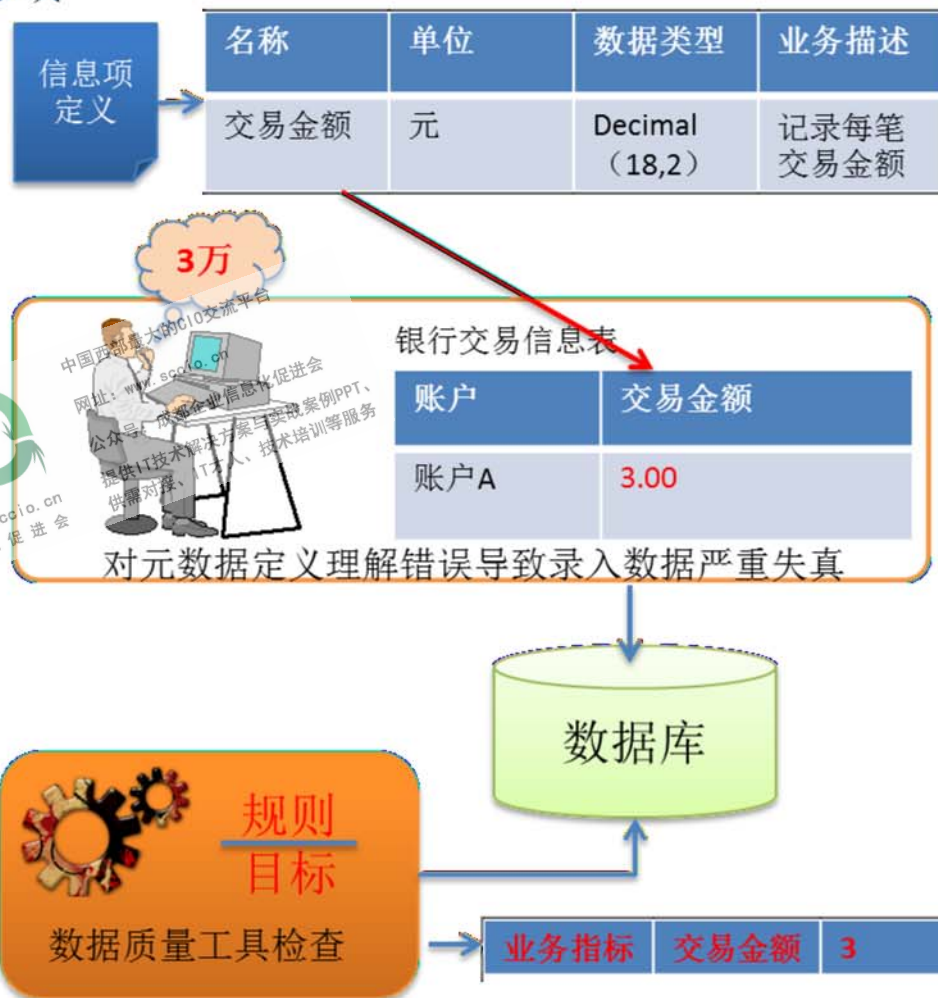
大数据质量管理工具介绍

大数据质量问题产生的原因分析—信息域

由于对大数据本身的描述、理解及其度量标准的偏差而造成的大数据质量问题。产生这类大数据质量问题的原因主要有：元大数据描述及理解错误、大数据度量的各种性质得不到保证、变化频度不恰当等。

元大数据描述及理解错误中的相关元大数据主要包括：

- 业务元大数据——主要包括业务描述、业务规则、业务术语、业务指标口径等；
- 技术元大数据——主要包括接口规范、执行顺序、依赖关系、ETL转换、大数据建模和工具等方面的内容。
- 大数据度量和变化频度提供了衡量大数据质量好坏的手段。大数据度量主要包括完整性、唯一性、一致性、准确性、合法性。变化频度主要包括业务系统大数据的变化周期和实体大数据的刷新周期。



大数据质量管理工具介绍

大数据质量问题产生的原因分析—技术域

由于具体大数据处理的各技术环节异常所造成的大数据质量问题，它产生的直接原因是技术实现上的某种缺陷。技术类大数据质量问题产生的环节主要包括：大数据创建、大数据获取、大数据传递、大数据装载、大数据使用、大数据维护等方面：

- 大数据创建质量问题主要包括：创建大数据默认值使用不当和大数据录入的校验规则不当，导致指标统计结果不一致、大数据无效、记录重复等；
- 大数据传递质量问题主要包括：接口大数据及时率低、接口大数据漏传、网络传输过程不可靠，如包丢失、文件传输方式错误、传输技术问题、协议使用不当导致的大数据不完整等；
- 大数据装载质量问题主要包括：大数据清洗算法、大数据转换算法和大数据加载算法的错误；



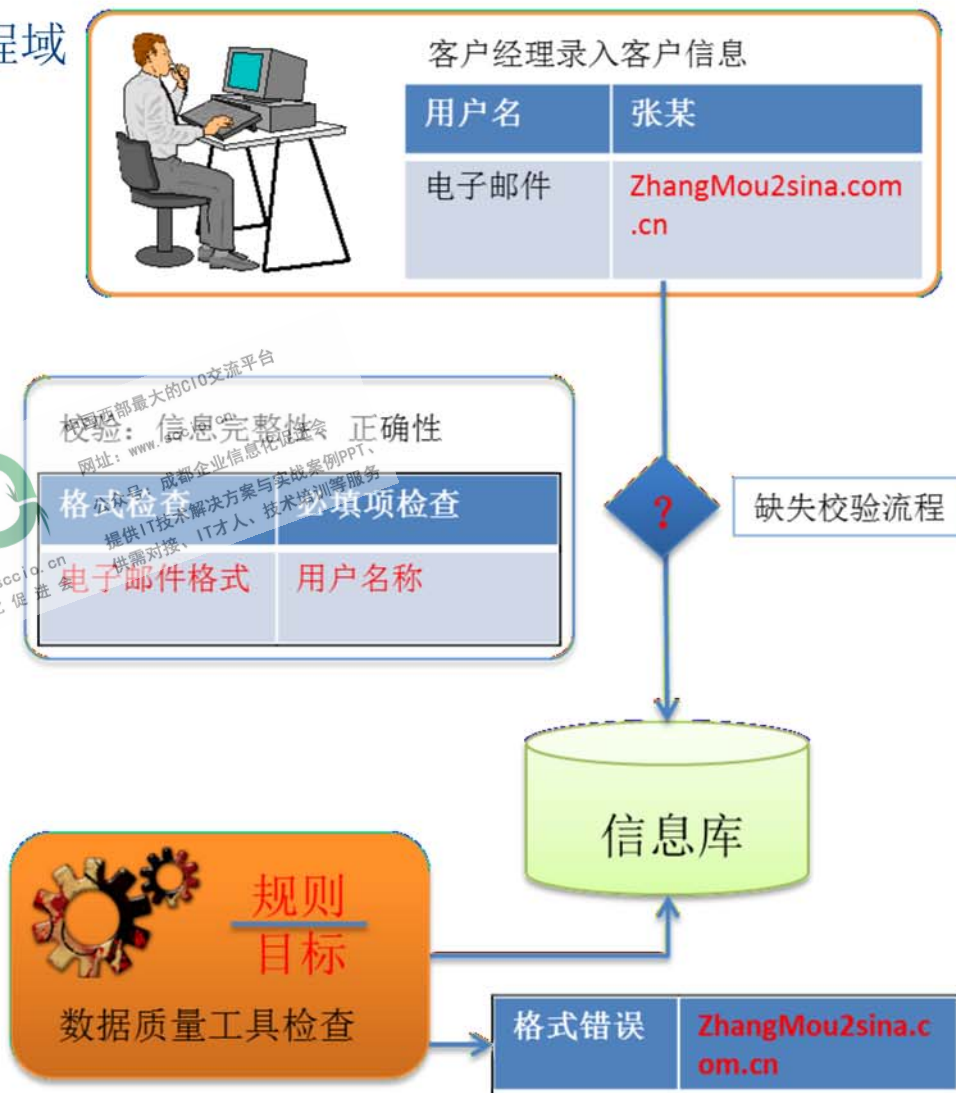
大数据质量管理工具介绍

大数据质量问题产生的原因分析—流程域

由于系统作业流程和人工操作流程设置不当造成的大数据质量问题，主要来源于系统大数据的创建流程、传递流程、装载流程、使用流程、维护流程等

各环节：

- 创建流程质量问题主要指操作员大数据录入时缺乏审核流程；
- 传递流程质量问题主要指通信流程沟通不畅；
- 装载流程质量问题主要指清洗流程缺乏/不当、调度流程逻辑错误、大数据加载流程逻辑错误及大数据转换流程逻辑错误；
- 使用流程质量问题主要指大数据使用流程缺乏流程管理；
- 维护流程质量问题主要指缺乏变更维护流程、缺乏错误大数据维护流程、缺乏大数据测试流程以及对人工后台调整大数据没有严格的流程监控；



大数据质量管理工具介绍

大数据质量问题产生的原因分析—管理域

由于人员素质及管理机制方面的原因造成的大数据质量问题如：

- 大数据数据库设计原则不严谨，大数据使用不规范导致的业务大数据重复，大数据不一致。
- 人员培训所产生的质量问题主要指对大数据质量相关人员缺少长期培训计划。
- 没有建立管理大数据质量的专门机构，出现大数据质量问题后无专人负责
- 没有明确的大数据质量目标；
- 缺少管理大数据质量的管理办法等。



规定：数据记录不能使用数据库代理键做主键，必须要有业务主键

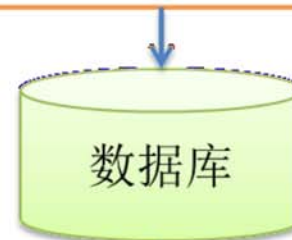
缺失管理约束将导致数据重复

中国西部最大的CIO交流平台
网址: www.sccio.cn
公众号: 成都企业信息化促进会
提供IT技术解决方案与实战案例PPT、供需对接、IT人才技术培训等服务

重复提交

银行交易信息表

数据库代理键PK	账户PK	客户名PK
UUID1	账户1	张三
UUID2	账户1	张三

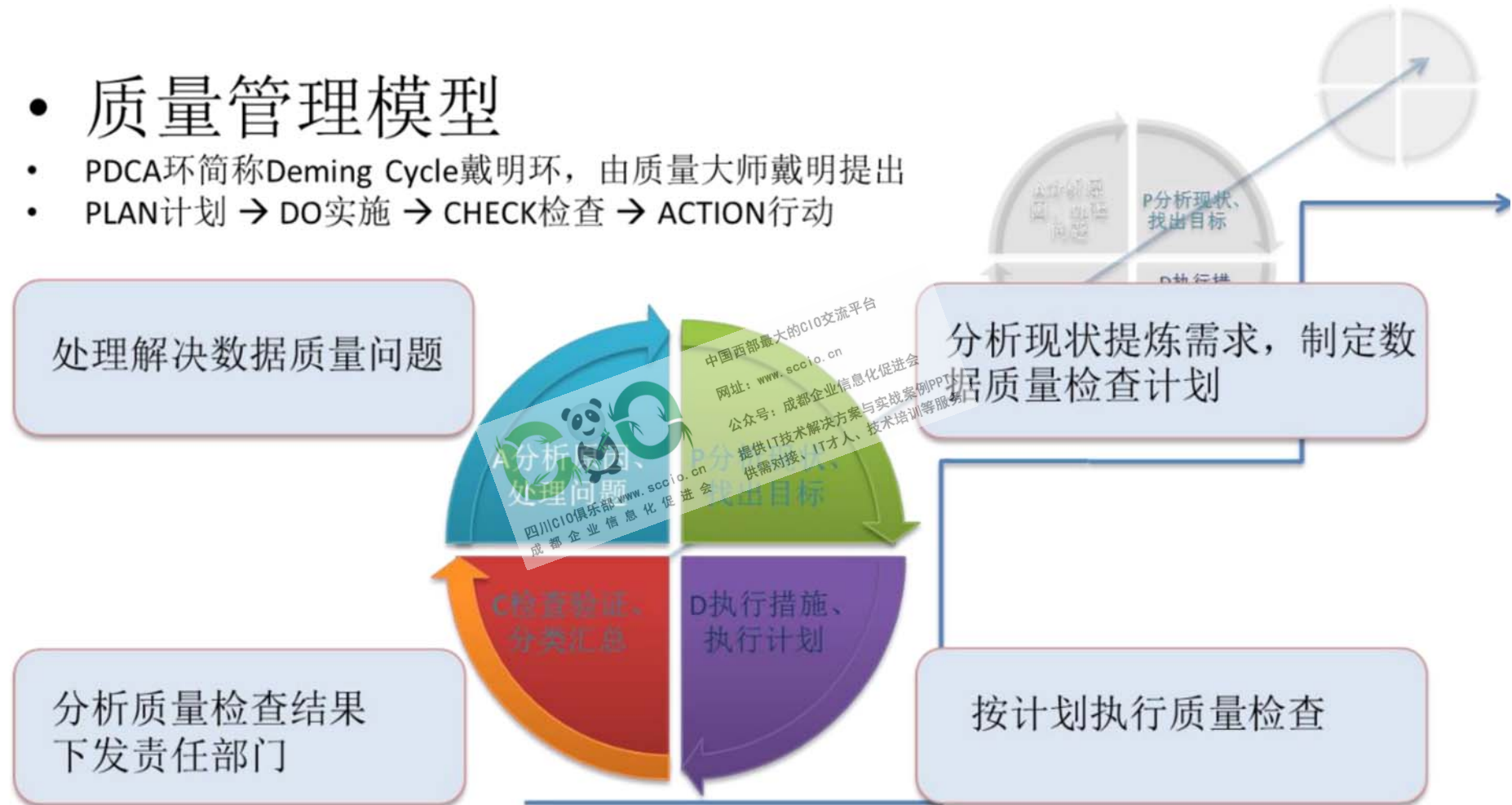


重复记录	账户1	张三
------	-----	----

大数据质量管理工具介绍

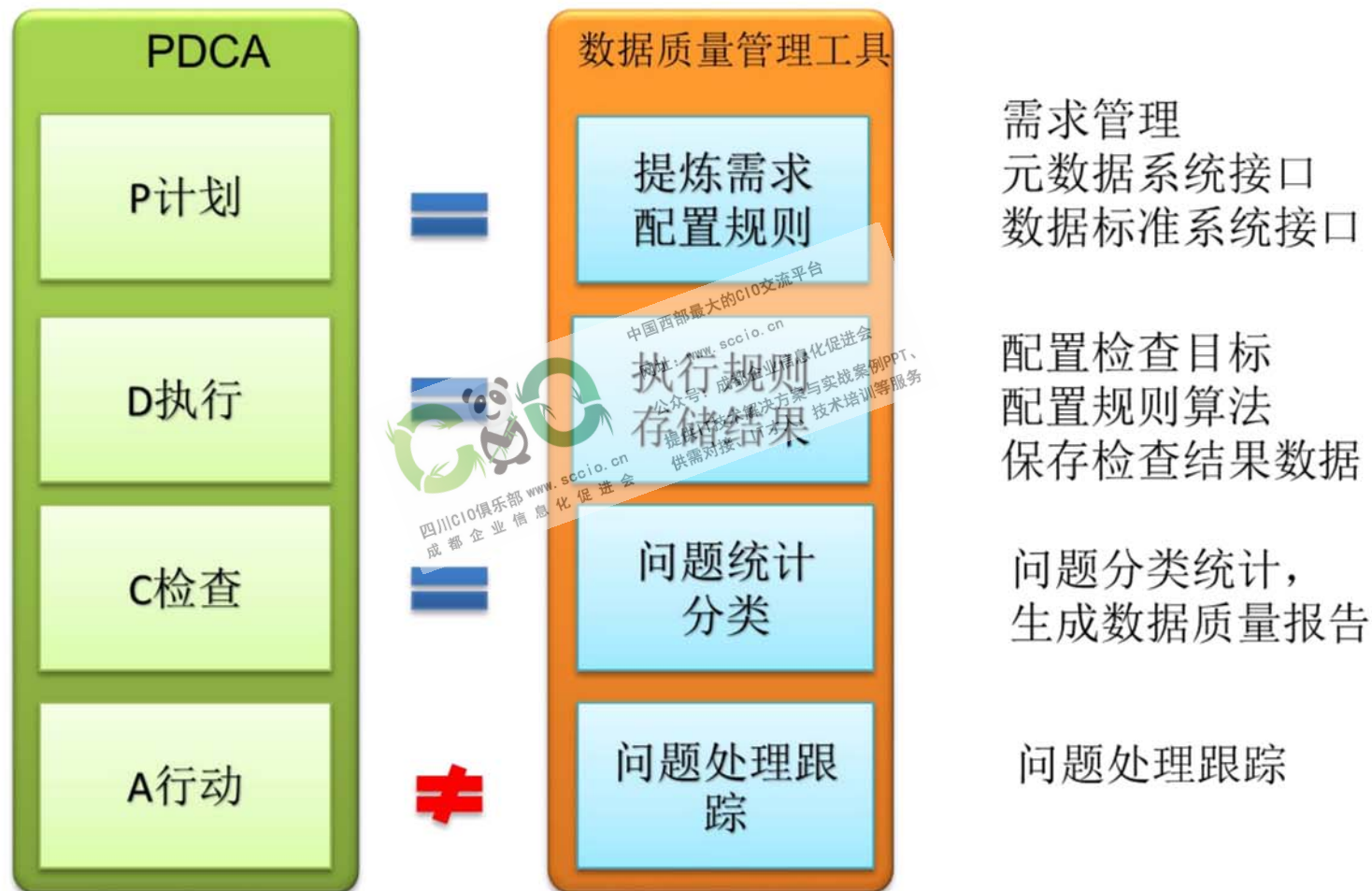
- 质量管理模型

- PDCA环简称Deming Cycle戴明环，由质量大师戴明提出
- PLAN计划 → DO实施 → CHECK检查 → ACTION行动



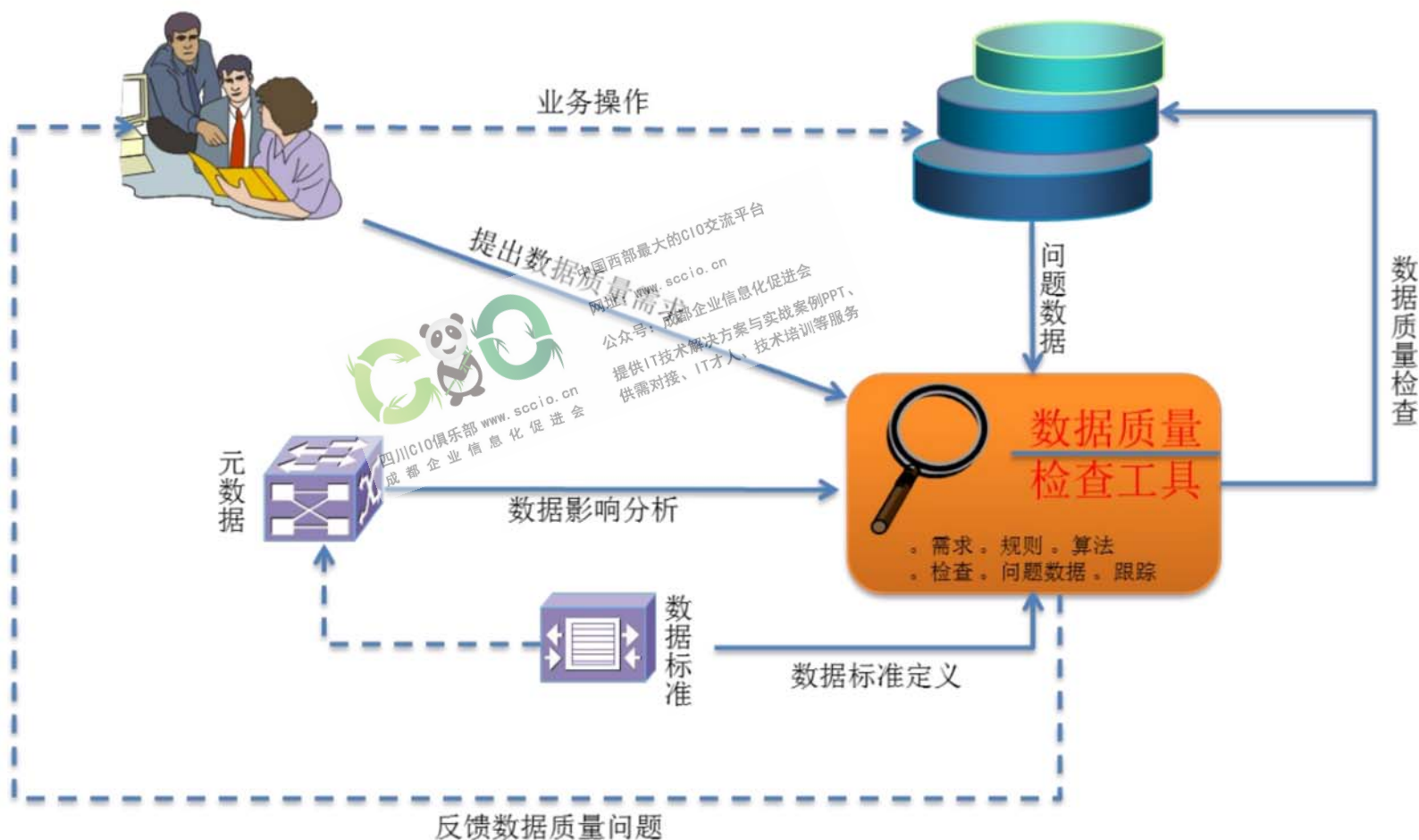
没有形成这个闭环，就不会有质量的改进

质量管理模型和功能匹配



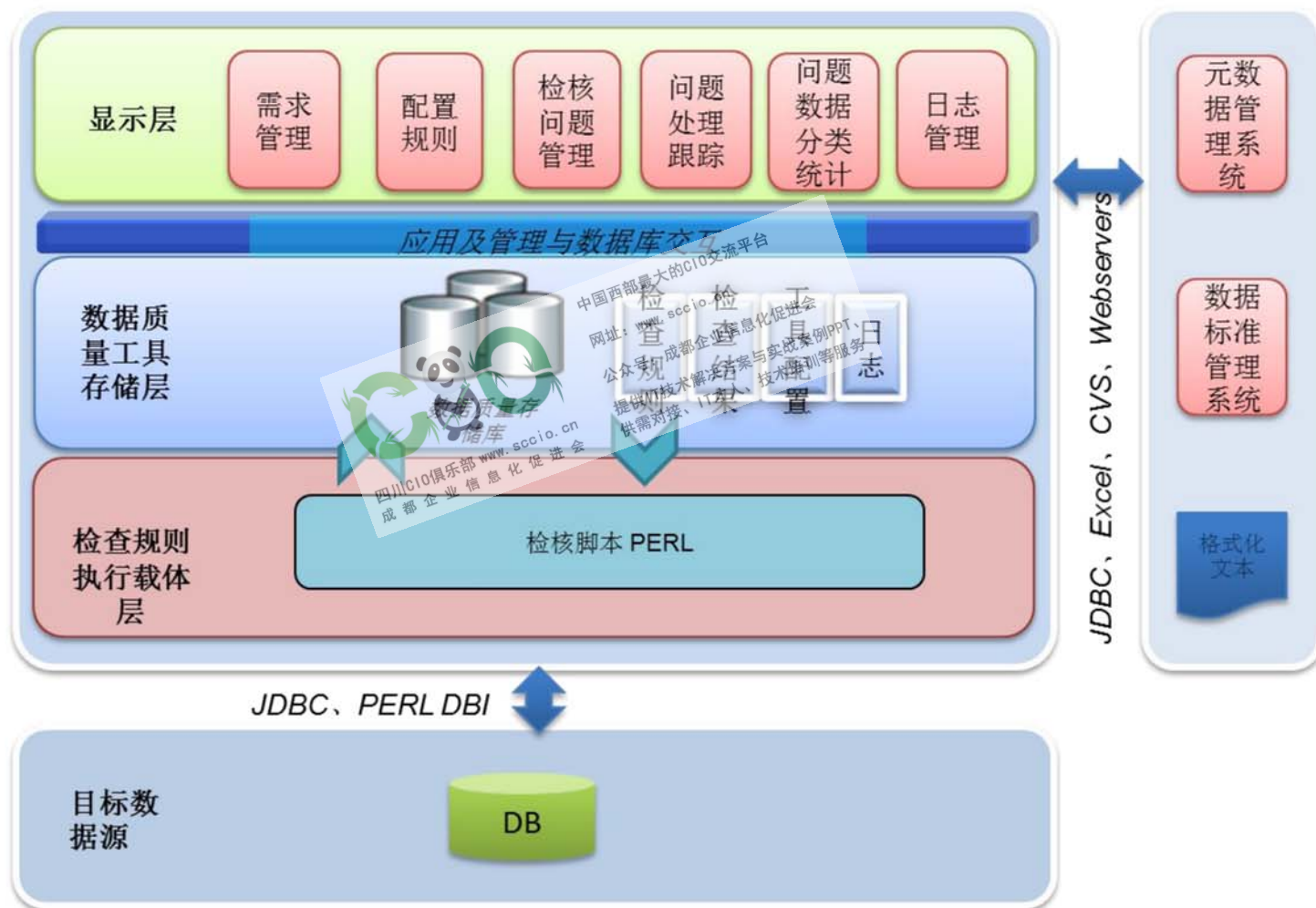
大数据质量管理工具介绍

数据质量管理工具运行逻辑图



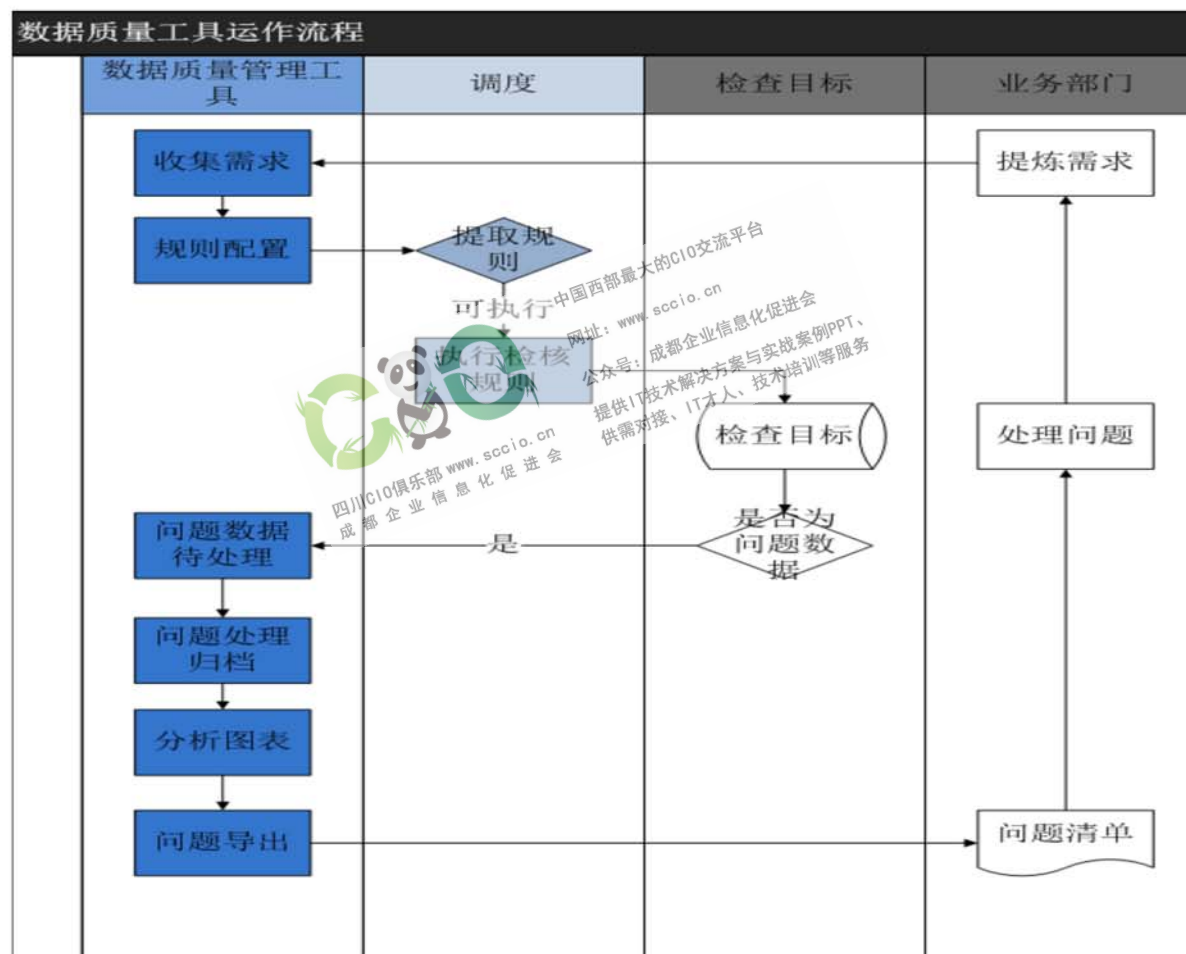
大数据质量管理工具介绍

数据质量管理工具架构图



大数据质量管理工具介绍

• 大数据质量管理工具流程图



大数据质量管理工具介绍

大数据质量管理开发功能清单

功能一级分类	功能二级分类	说明
首页	首页统计分析	支持最近一天以及最近一周的质量情况统计分析，支持钻取；支持个性化首页定制；
规则管理	规则配置	规则依赖的需求，需要执行的大数据标准，业务描述、规则类型（完整性，一致性，时效性，正确性）
	规则浏览	规则清单浏览
检查点管理	检查点配置	配置检查目标及规则，技术属性（唯一，非空，值域，代码，外键，拉链）
	检查点浏览	检查点清单浏览
	检查点执行状态监控	检查点手动执行、状态监控
	检查点执行历史查询	检查点运行历史查询，执行情况浏览
问题跟踪分析管理	问题浏览	检查点执行后的问题大数据（综合问题查询）
	问题发布	将问题提交给处理人，包括处理人信息、预期处理完成时间、提醒方式等，支持已发布的问题大数据浏览
	问题处理	用于关闭问题。记录问题处理过程及方法。
	问题归档	处理完成后的问题的归档，支持已归档的问题大数据浏览
	问题跟踪监控	反映问题严重级别，预期处理时间，预定处理部门或责任人，触发通知等功能。
	问题跟踪设置	设置问题大数据处理日历。问题严重级别。
大数据质量分析	大数据质量全局统计	全局统计问题报告，系统维度，大数据生成周期，处理情况
	单表问题大数据统计	单个检查目标问题大数据类型统计图。
	单检查点趋势分析	单个检查点趋势图
	问题分析报告下载	问题分析报告下载
权限管理	用户管理	用户录入，修改。密码管理
	角色管理	创建角色、用户分配角色
	权限配置	系统功能权限分配
菜单管理	菜单管理	系统菜单新增、修改。
日志管理	系统监控日志	用户操作系统的记录，包括功能点，请求参数
	检查执行日志	检查点执行时间，返回状态码，错误信息。
检查脚本	检核脚本（PERL模板）	后台调用检查配置信息的Perl脚本。



中国西部最大的CIO交流
网址: www.sccio.cn
公众号: 成都企业信息化促进会
提供IT技术解决方案、技术咨询、
供需对接、IT人才、技术交流等

大数据质量与元大数据的关系

- ❑ 元大数据侧重于展现表结构化的信息，大数据质量侧重展现表中大数据存在的问题；
- ❑ 元大数据可以展现表上与大数据质量相关的信息；
- ❑ 大数据质量可以利用元大数据的分布拓扑结构图信息（大数据地图）展现大数据质量点、线、面的与质量相关的分布情况；
- ❑ 大数据质量发现问题时可以查看相关表结构

