

**BDTC**

2014 中国大数据技术大会

BIG DATA TECHNOLOGY CONFERENCE

暨第二届CCF大数据学术会议

# 大数据时代的掘金术

——数据资产管理探讨

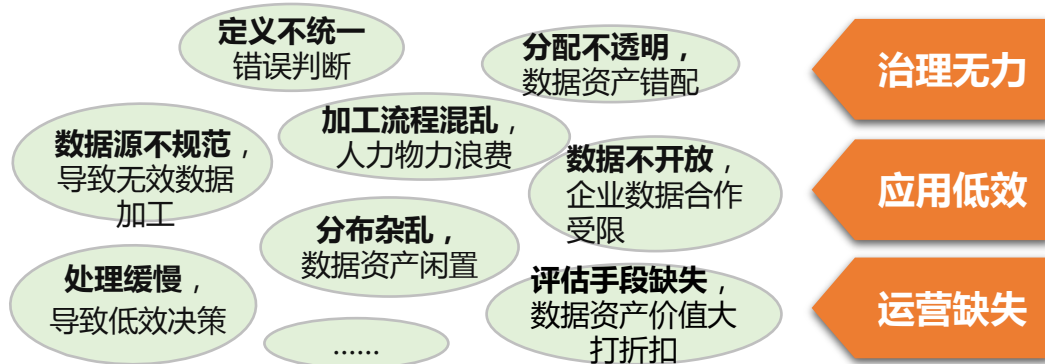
# 尽管“数据是资产”概念已经广为人知，但“如何管理数据资产”仍然缺少成熟理论以及工具手段

## 什么是数据资产？



数据资产是企业及组织拥有或控制，能带来未来经济利益的数据资源。

## 存在什么问题？



## 需求发现

数据资产管理是企业或组织采取的各种管理活动，用以保证数据资产的安全完整，合理配置和有效利用，从而提高带来的经济效益，保障和促进各项事业发展。该领域是大数据时代企业布局竞争的核心，也是目前市场空白。

## 数据资产管理包括哪些核心内容？

### 数据资产治理

让企业数据更加准确、一致、完整、安全，降低IT成本。

### 数据资产应用

使得企业数据的使用过程更为人性、快捷、智能，从而提升管理决策水平。

### 数据资产运营

支持企业数据资产的分发、开放、交易等数据嫁接的实现，从而促进数据资产的价值实现。

# 为什么传统数据管理方式不适合数据资产管理要求？

## 传统数据管理方式

外部性管理，依赖管理力度和执行自律，成难毁易。

元数据

数据  
稽核

管理  
制度

从**范围**来看，



非结构化数据、内外部数据混搭、云化处理等都会冲击传统管理模式

**挑战1**

从**形式**来看，



数据加工的复杂度和速度要求越来越高，也对传统管理效率提出挑战

**挑战2**

从**内涵**来看，



资产验证

数据整合

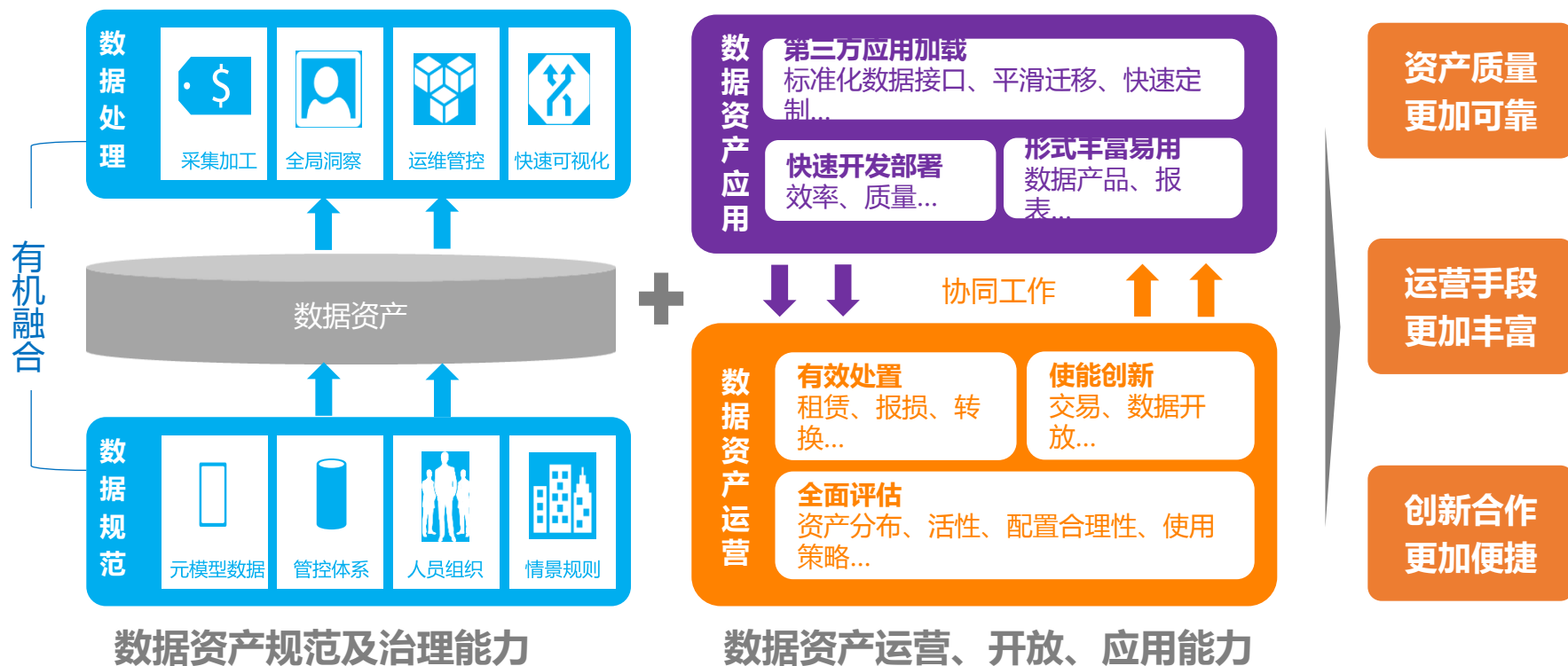
交易保障

数据的交换、转让、租赁、交易等各种创新模式，也要求新的管理手段

**挑战3**

# 亚信提倡建立一体化全流程的数据资产管理体系

数据资产管理体系核心在于有效解决对数据资产进行管理的实践性问题，既帮助企业合理评估、规范和治理企业信息资产，又可以挖掘和发挥数据资产价值并促进持续增值，并符合大数据的跨行业合作趋势



# 数据资产管理平台在企业IT系统中的定位

定位于应用支撑层，在数据治理基础上，实现资产规划、加工、评估、运营等功能





# 关键特性介绍：完善的数据治理与管控（1/5）

## 1 建立标准体系

数据标准是数据资产管理的基础，需要对管理对象、管理要求、管理手段、管理流程等进行规范，从而成为海尔相关系统和部门统一遵循的标准。

## 2 形成信息地图

建立全集团的IT的信息地图，通过自动化的多源头元数据采集，自动分析汇总，形成完整的企业数据地图，使用户能够从全局视角审查企业整体数据状况。

## 3 具备管控手段

针对数据生命周期各个阶段的不同特性，提供各种监控、管理工具，将可能出现的系统运行出错或数据异常变化进行修正或告警，避免出现更大的损失

## 建立起可管可信的数据资产治理体系

## 4 实现影响分析

实现数据来源的追溯，能够方便内部管理、审计或外部监管的需求追溯业务指标、报表的数据来源和加工过程，即能方便的找到想要的的数据以及这个数据与其他数据的传递关系和业务逻辑关系。。

## 5 促进数据协同

实现跨平台的元数据管理，具备数据管控统一功能平台，增强应用的协同管理能力，能够展示出数据之间的关系，从而促进不同阶段的数据形成协同关系，以及闭环加工流程，确保数据可靠性。

## 6 持续质量改进

数据质量体系需要通过实践和规划的相互促进，不断完善改进，为此，需要确保数据架构合理，条理清晰，过程可控，知识积累传承，并通过监控和审计不断促进质量水平的持续提升。

## 关键特性介绍：完善的数据治理与管控（2/5）

- 通过对数据、应用、系统综合管理，构建标准化、流程化、自动化、一体化的数据管理体系
- 确保数据架构合理，条理清晰，过程可控，知识积累传承。



**基础分类标准**

**数据权限规则**

**元模型标准**

**数据分类标准**

**命名规范要求**

**存储规则**



## 关键特性介绍：完善的数据治理与管控（3/5）

DACP提供了完整全面数据治理与管控功能体系，可以帮助企业实现数据资源的条理化、脉络化，成为数据资产化管理的重要基础

### 专业数据管理

- 数据标准化制定
- 数据架构管理
- 主数据管理
- 指标数据管理
- 代码标准化管理
- 规则稽核评估

### 高效数据处理

- 数据采集
- 数据加工
- 数据分发
- 数据共享
- 敏感数据管理

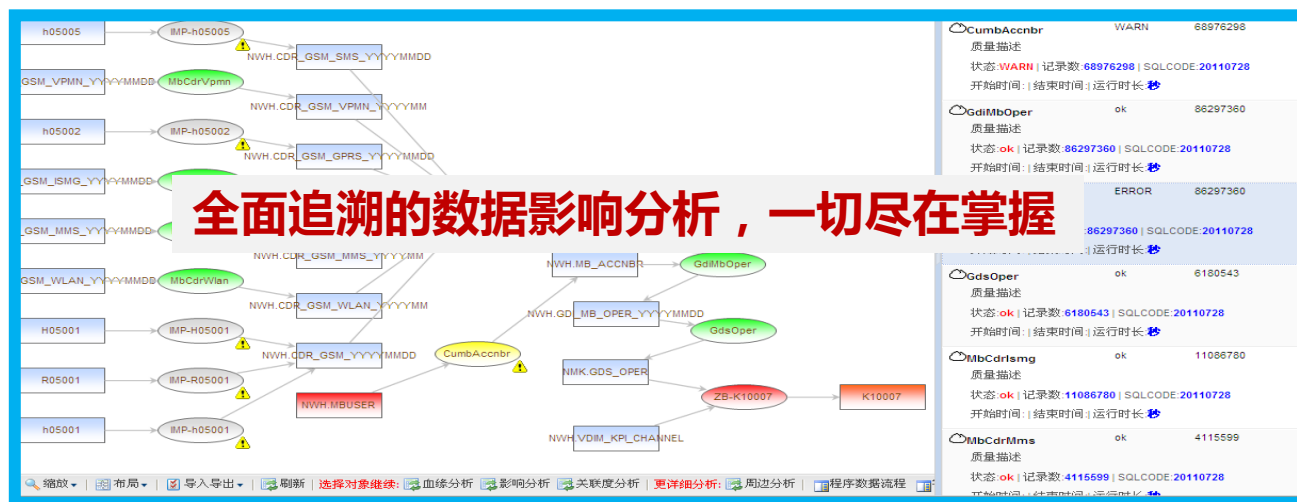
### 持续质量改进

- 质量规则管理
- 问题定位分析
- 影响范围分析
- 问题知识库
- 质量标准定位
- 数据标准化管理

### 快速响应

- 可视化开发管理
- 需求分析
- 变更分析
- 知识积累
- 自助分析
- 数据可视化

# 关键特性介绍：完善的数据治理与管控（4/5）



系统提供完整细致的血缘分析，对问题的节点进行回溯，分析其处理路径上可能存在的问题以及相关影响范围

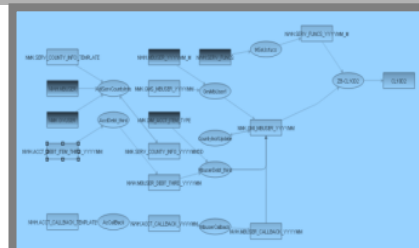
## 接口

接口名称	接口类型	接口地址	接口描述	接口状态
h05005	接口	h05005	h05005	h05005
h05002	接口	h05002	h05002	h05002
h05001	接口	h05001	h05001	h05001
R05001	接口	R05001	R05001	R05001

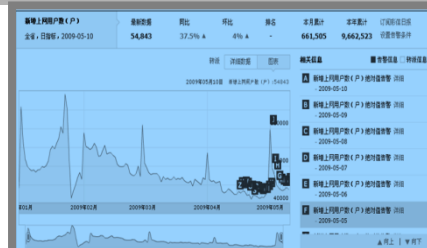
## 数据表

数据表名称	数据表类型	数据表地址	数据表描述	数据表状态
h05005	数据表	h05005	h05005	h05005
h05002	数据表	h05002	h05002	h05002
h05001	数据表	h05001	h05001	h05001
R05001	数据表	R05001	R05001	R05001

## 处理程序



## 应用



# 关键特性介绍：完善的数据治理与管控（5/5）

## 一站式统一运维监控

- 为运维部门提供了一个中央管理点，使得运维人员可以紧密有效地对系统上发生的事件进行控制，为分布式环境创建一个“任务控制”中心。
- 对收集到报警信息及时触发各种动作，可通过邮件、短信、语音等方式提醒运维人员
- 对日常监控数据的分析，也可以帮助运维人员分析出系统中存在的性能瓶颈，以便采取适当的解决措施对系统进行优化或扩展。



# 关键特性介绍：高效的数据资产应用（1/4）

通过数据管理来提升开发效率，而加强开发管控反过来也促进了数据质量有效提升

全过程的开发管理，提升  
执行效率，确保数据质量

- 需求
- 设计
- 开发
- 调试
- 测试
- 部署、升级
- 文档生成



# 关键特性介绍：高效的数据资产应用（2/4）

## 丰富的组件库快速帮助实现数据资产应用

### UI组件

提供方便灵活的组合方式，并能和数据组件进行绑定



### 社区组件

以组件的形式保证应用在其内部的事务控制动作



### 规则组件

封装了特定业务逻辑，有明确的输入和输出，保证业务规则的实现



### 分析导航组件

提供数据获取能力，对获取的数据需要提供明确的数据指标、指标口径、数据范围、数据时效性等



### 安全服务组件

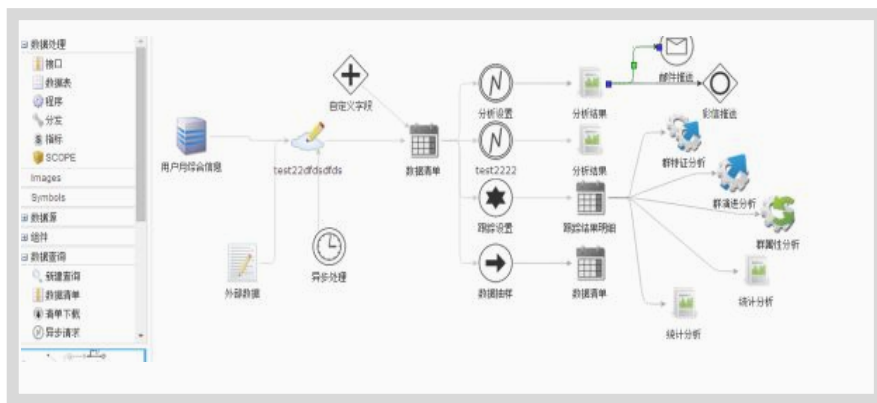
负责BI Store 应用使用日志的记录，以及与经营分析系统安全模块交互进行客户端使用权限控制



# 关键特性介绍：高效的数据资产应用（3/4）

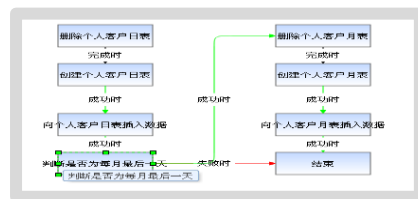
体验良好、可靠高效的设计开发过程

可视化的拖拽式开发

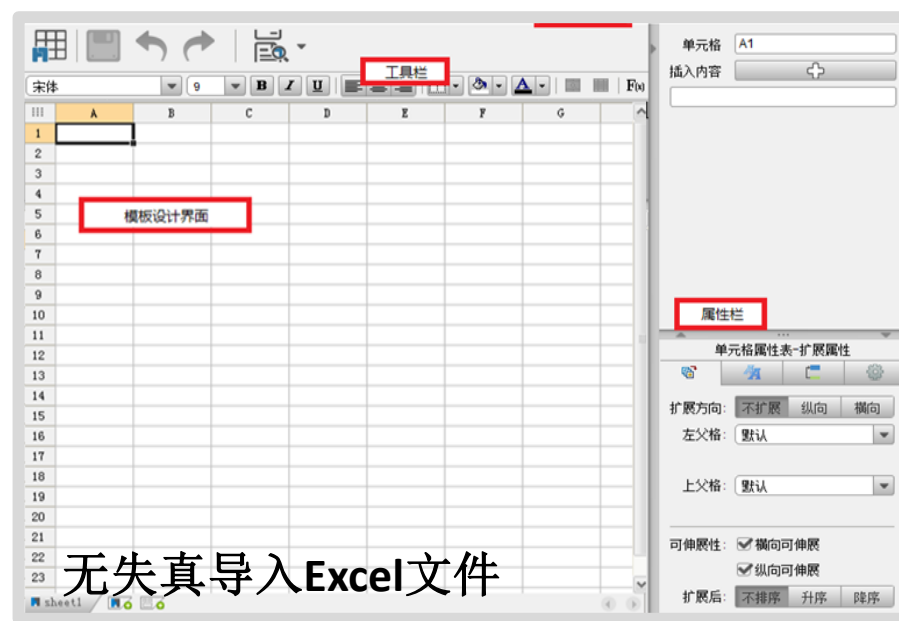


增加 批量导入 类似创建 脚本 报表脚本 比较 导出 刷新 保存

序号	字段名	数据类型	长度	备注
1	GROUP_ID	客户群ID	VARCHAR(21)	1
2	GROUP_NAME	客户群名称	VARCHAR(21)	0
3	GROUP_TYPE	客户群类型	VARCHAR(21)	0
4	GROUP_DESC	客户群描述	VARCHAR(21)	0
5	GROUP_CHARA	客户群特征	VARCHAR(21)	0
6	REGION_CODE	归属地区	VARCHAR(21)	0
7	CUST_NUM	客户数	VARCHAR(21)	0
8	VALID_FLAG	是否有效	VARCHAR(21)	0
9	CUST_LIVE_FLAG	是否活跃	VARCHAR(21)	0



类Excel的报表设计过程





## 关键特性介绍：高效的数据资产应用（4/4）

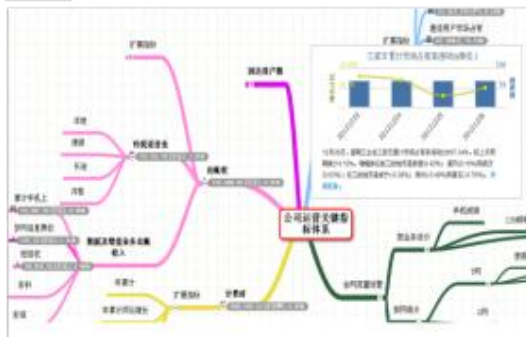
### 丰富美观、实用大方的数据可视化能力

指标墙

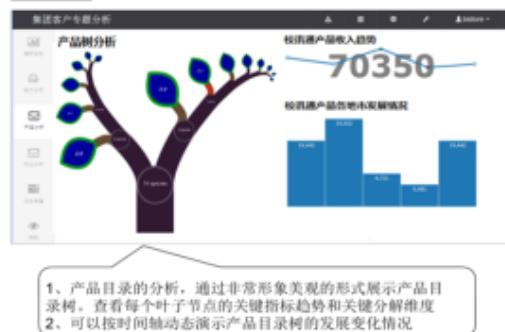
- 1、查看指标列表，定制自己关注的指标
- 2、借鉴web2.0体验设计，“就地查看”，在指标墙，双击指标直接查看最关键时空情况



脑图分析



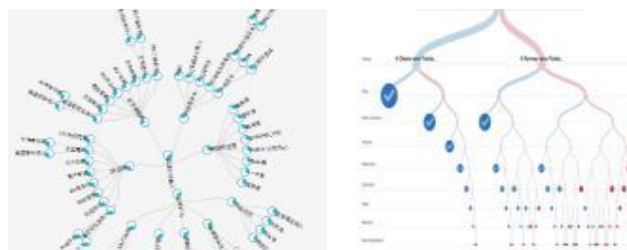
产品树分析



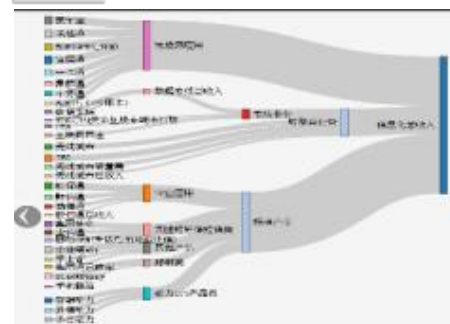
GIS可视化分析



关联分析

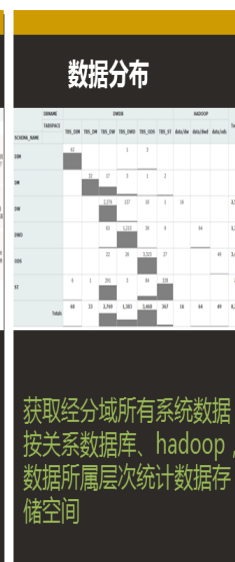
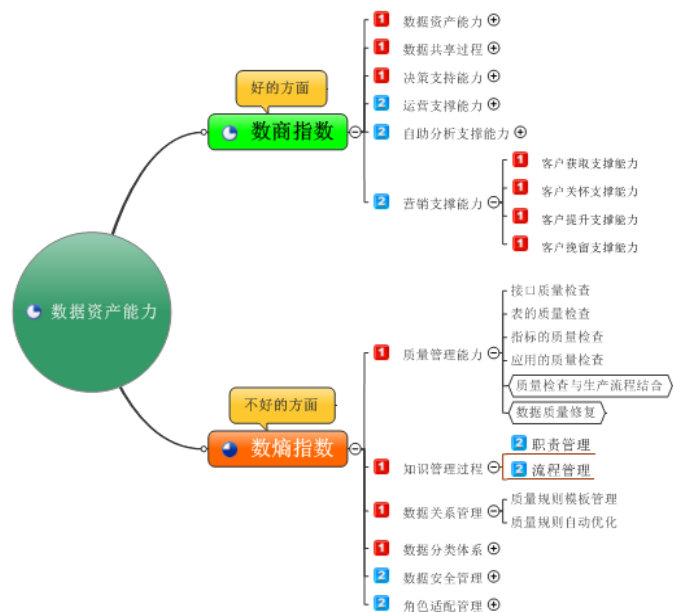


数据流向



# 关键特性介绍：创新的数据资产运营（1/4）

根据多年的经验积累，总结了从数据到资产的评估体系，可以帮助企业全方位的诊断发现问题，并提供相应的策略来企业提高数据的健康度



# 关键特性介绍：创新的数据资产运营（2/4）

## 支持数据资产多团队使用的统一管理



要实现不同团队的统一管理，需要解决如下问题

- 如何实现不同开发商输出结果的统一化标准化？包括表命名的规范、数据字典
- 知识的发布和共享，数据共享
- 如何实现统一的运行监控？统一调度管理
- 如何数据权限控制、运行资源的控制
- 如何实现统一的数据管理？
- ....



- **支持多开发团队的管理功能。**包括数据权限分配，CPU、I/O资源管理
- **支持不同数据平台统一开发。**通过抽象API屏蔽平台的差异性，实现不同数据平台的统一开发。
- **统一元数据管理。**每个开发团队输出到不同的开发目录。内容包括现有的数据字典、业务口径、程序代码等
- **统一的标准与规范。**不管哪个开发小组开发的内容必须满足这些基本的标准。
- **共享与授权。**数据安全可控的范围内，促进数据的复用与共享。
- **统一运行调度管理。**多团队开发输出处理任务彼此之间存在复杂逻辑关系。统一调度管理能够稳定智能去做任务运行编排

# 关键特性介绍：创新的数据资产运营（3/4）

企业应用商店



- ✓ 统一运行环境，主机，数据库等
- ✓ 打通与各系统的接口。
- ✓ 统一数据中心
- ✓ 统一的开发工具
- ✓ 提供可二次开发的SDK包
- ✓ 应用发布评估管理
- ✓ 运行监控
- ✓ 下线管理
- ✓ 应用类型包括报表，信息推送，手机积分，手机客户端等

通过标准化数据服务，搭建企业的应用开放平台，促进应用的百花齐放

运行基础平台

- ✦ 运行资源的稳定性
- ✦ 应用上线下线管理
- ✦ 应用共享管理
- ✦ 信息安全管理
- ✦ 统一数据访问接口
- ✦ 统一应用运行机制

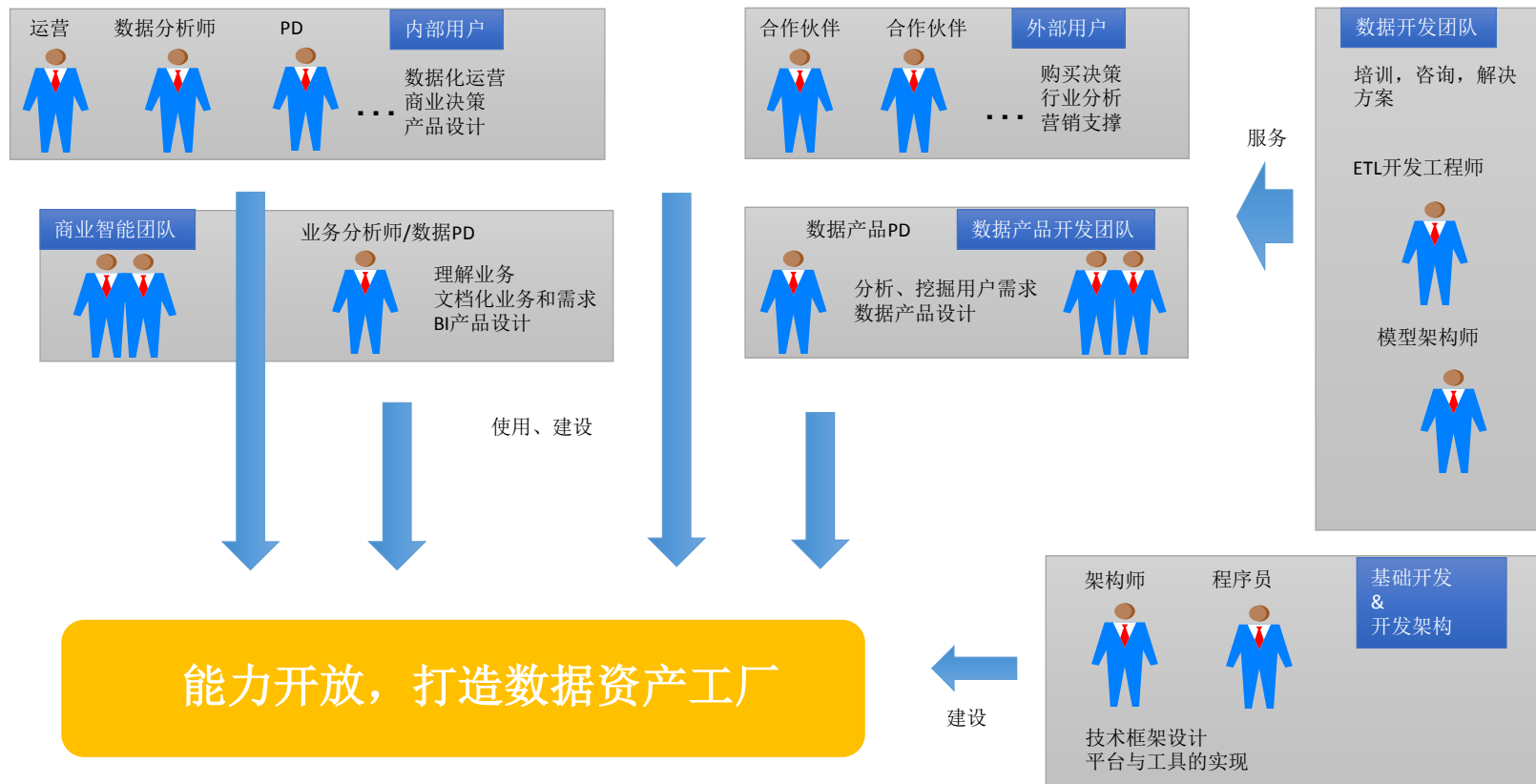
数据标准化服务

- ✦ 数据处理开发；
- ✦ 数据高速访问提供
- ✦ 数据安全控制
- ✦ 数据沙箱提供
- ✦ 32类数据开放

基于Api二次开发

- ✦ 封装好的sdk API进行开发
- ✦ 信息推送API支持
- ✦ 展示组件支持
- ✦ 订阅中心

# 关键特性介绍：创新的数据资产运营（4/4）

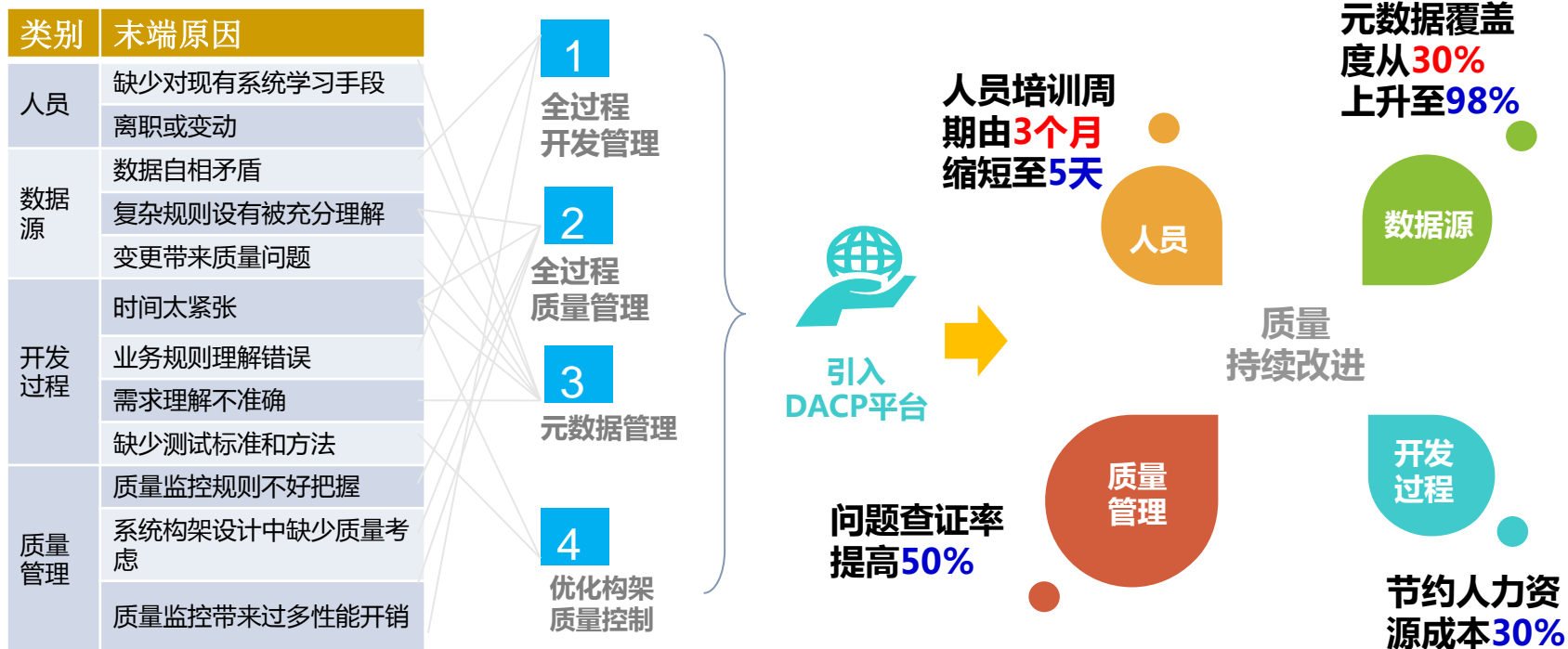




# 典型案例1：助力数据治理、提升工作效率

客户现状：

- 1、BI系统经过12年以上建设，超过10万张数据表、1600个以上应用，容量超过100T
- 2、存在多厂家合作开发，数据质量管理主要依靠管理制度和事后稽核





## 典型案例2：高效建设大数据应用系统，管理海量数据

### 客户现状：

- 1、客户计划采取Hadoop+DB2 混搭式结构实现经分域系统重构、优化。
- 2、2040个接口、功能模块，1058个模块,2700多个指标,2306个数据表重构,180个一经上报，参与团队多、周期长。

构建时间	2013-6-1			
管理节点	2台	32C	64G	300G*2 2T*6
横云平台	2台	32C	64G	300G*4
计算节点	27台	32C(16C)	64G	300G*2 2T*6
	8台	48C	64G	300G*2 1T*6
存储能力	310T(已使用190T)			
计算能力	Map	620		
	Reduce	310		

效果：

- ✦释放了仓库的高端存储**25T**；
- ✦释放了仓库**50%**的计算能力；
- ✦通过新云平台将传统仓库的详单数据的存储周期从**60天提升到210天**；

云平台数据说明						
数据类型	大类数	计划存储周期	实际存储周期	是否压缩	单周期数据量	数据总量
ODS	17	60	100天	否	0.5	90
DWD	17	210	90天	是	0.4	30
DW	42	210	90天	是	0.2	14
DM	10	370	90天	是	0.02	1
OutPut	50	1	1	是	0.3	0.1
其它						10

后续将690的动态划分库的历史结果划分明细导入新的仓库后将释放整个690动态划分库。

存储了3份，合计190T

效率提升

跨平台，跨系统数据抽取分发。数据处理效率提升30%

可视化的开发平台，建设周期缩短20%

高效、灵活的报表展现工具，二次开发效率提升40%

能力增强

计算速度提升120%，存储能力提升5倍，为过渡到企业级大数据中心奠定基础

DACP投入使用还释放了仓库的高端存储25T；释放了仓库额外50%的计算能力；同时通过新平台将传统仓库的详单数据的存储周期从60天提升到210天。

成本降低

整体性能差不多的情况下，采用DACP投资降低20%左右；

## 典型案例3：基于应用平台，实现自主取数分析

客户现状：

- 1、业务部门需要开展营销活动或进行业务分析，则以工单形式要求IT部门临时统计或取数。
- 2、IT部门投入8个人负责对口，每个月完成临时统计300次以上，处理数据稽核90次以上。

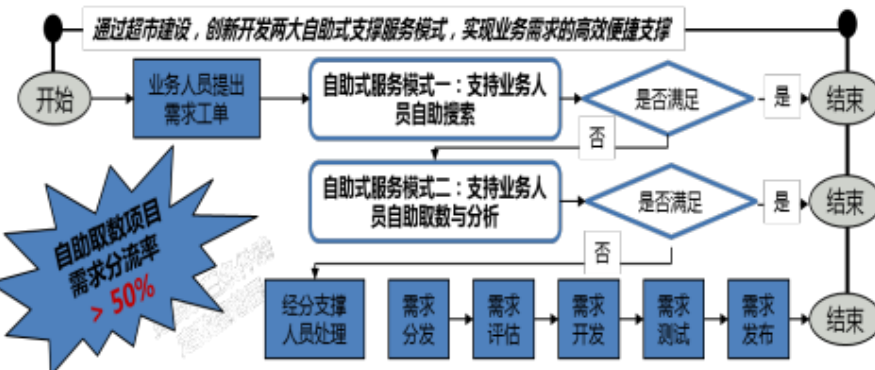
自助分析是以业务人员特别是面向一线人员为使用对象，以实现灵活、快捷的细粒度数据组织、提取及可视化的二次分析为核心目标，以易用、稳定、快捷、严谨为宗旨的一款工具类产品。



传统支撑模式：需求满足总耗时>5天，需求满足度<60%



自助式支撑新模式：需求满足总耗时<5分钟，需求满足度>80%



# 数据资产管理平台，不仅是技术工具，也是重要的运营管理手段！



## 良好界面体验

全Web界面的在线操作  
提供丰富的可视化组件，能够  
很好地帮助信息展现、问题定  
位和决策支持



## 完善规范标准

凝聚了电信级数据管理规范  
和实践经验，形成一整套管理办  
法和信息体系

## 支持第三方应用集成



### 一站式管理：

- 面向业务人员提供快速取数、自助分析、门户定制等功能；
- 面向技术人员提供可视化运维、自动化处理等工具；
- 面向管理人员提供各种评估告警以及决策支持手段；



## 使能数据开放

支持多租户管理、数据开放平  
台、数据合作加工模式等数据  
交互模式，并能够很好地保证  
数据使用过程中的隐私安全；



## 智能化过程控制

能够根据数据资产实际使用过  
程进行智能化分析，并动态调  
整管理过程中的规则参数；

数据处理及可视化框架



## 数据治理



企业级大数据中心

**BDTC**

2014 中国大数据技术大会

BIG DATA TECHNOLOGY CONFERENCE

暨第二届CCF大数据学术会议

谢谢聆听！