

# 数字化转型中的大数据治理架构

**1** 数字化时代大数据向服务化发展

**2** 大数据治理架构与关键技术分析

**3** 普元大数据治理架构实践

突破狭义互联网的限制，认识虚拟与现实的本质：  
虚拟时空的交易与现实时空的交付



**互联网**  
信息互通

网上银行、银证通、银保通

**移动互联网**  
时间自由 空间自由

手机银行、移动支付

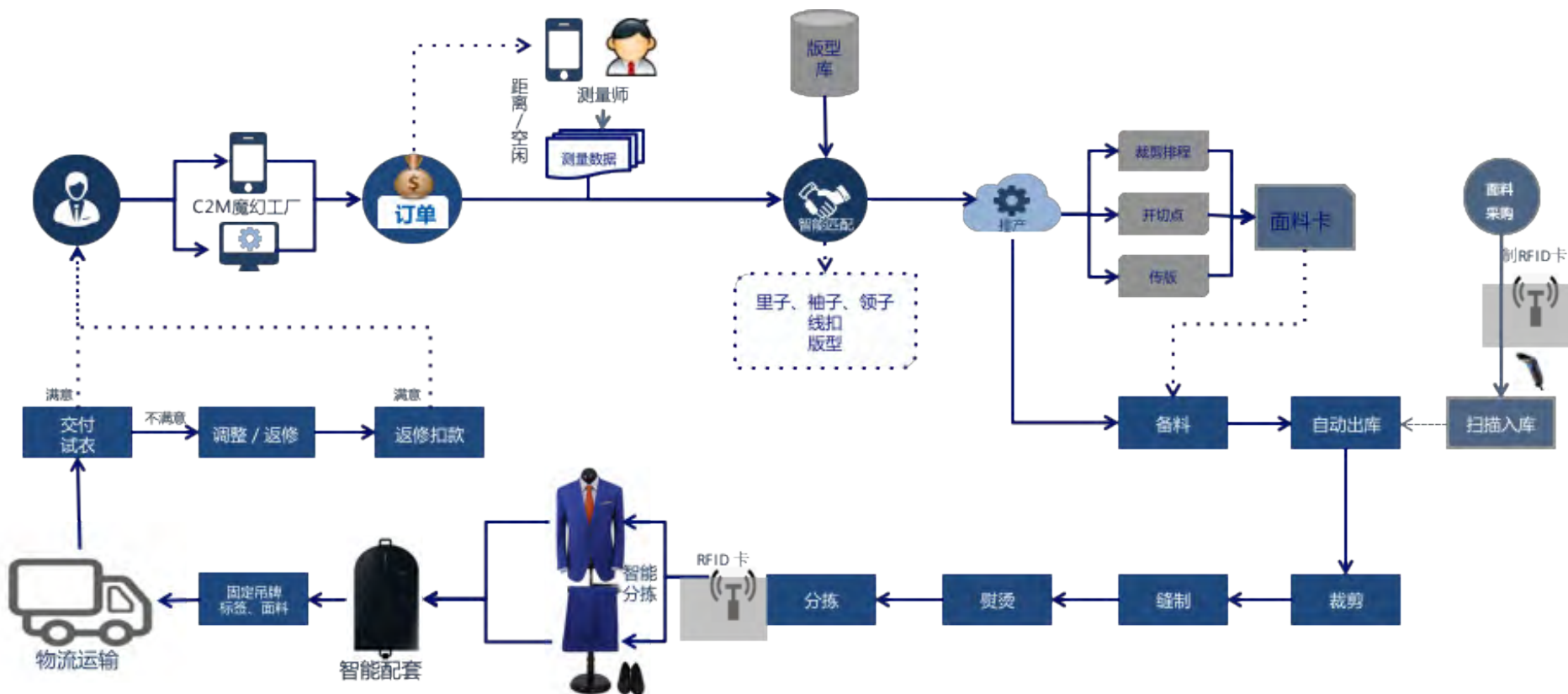
**互联网+**  
虚拟时空&现实时空

消费金融、工业4.0

**人、物、商业数字化连接**的商业模式



红领集团通过**业务创新**实现由客户需求直接驱动工厂的运作模式

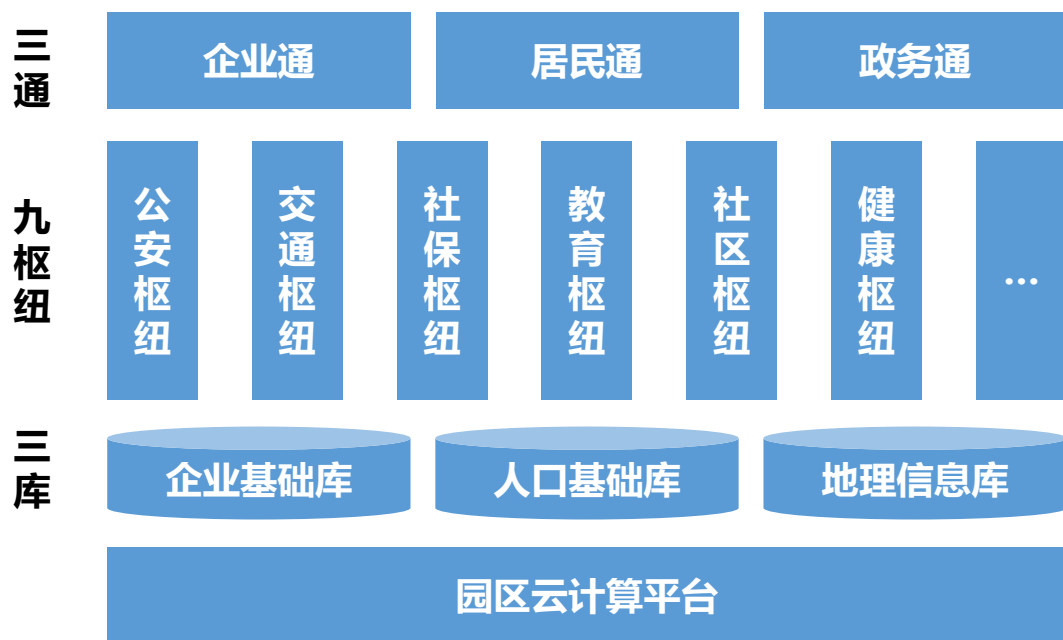




红领集团通过**业务创新**实现由客户需求直接驱动工厂的运作模式



苏州工业园区在信息化发展规划中提出了“**三库、三通、九枢纽**”的信息化总体发展框架。



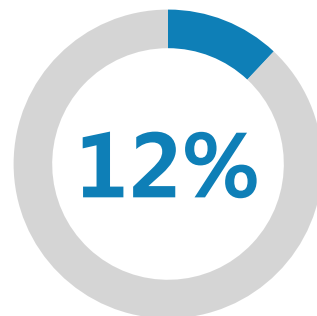
接入园区**31**个部门、涉及**85**个业务领域，**335**个数据分类，**3981**工作项。

先后完成了**公安、工商**等**15**家**人口、法人**信息高关联度局办的信息纳入政务资源云平台。

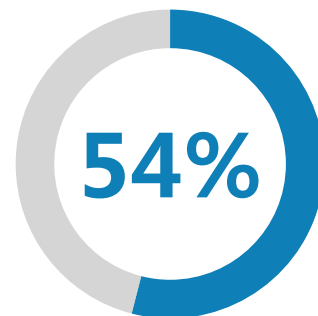
根据大数据咨询公司NewVantagePartners近期发布的调查报告，在接受调查的企业中，2012年仅有**12%**任命了首席数据官，而2016年这一比例上升到**54%**



2008-2009年世界金融危机，  
第一位CDO诞生



**2012年**

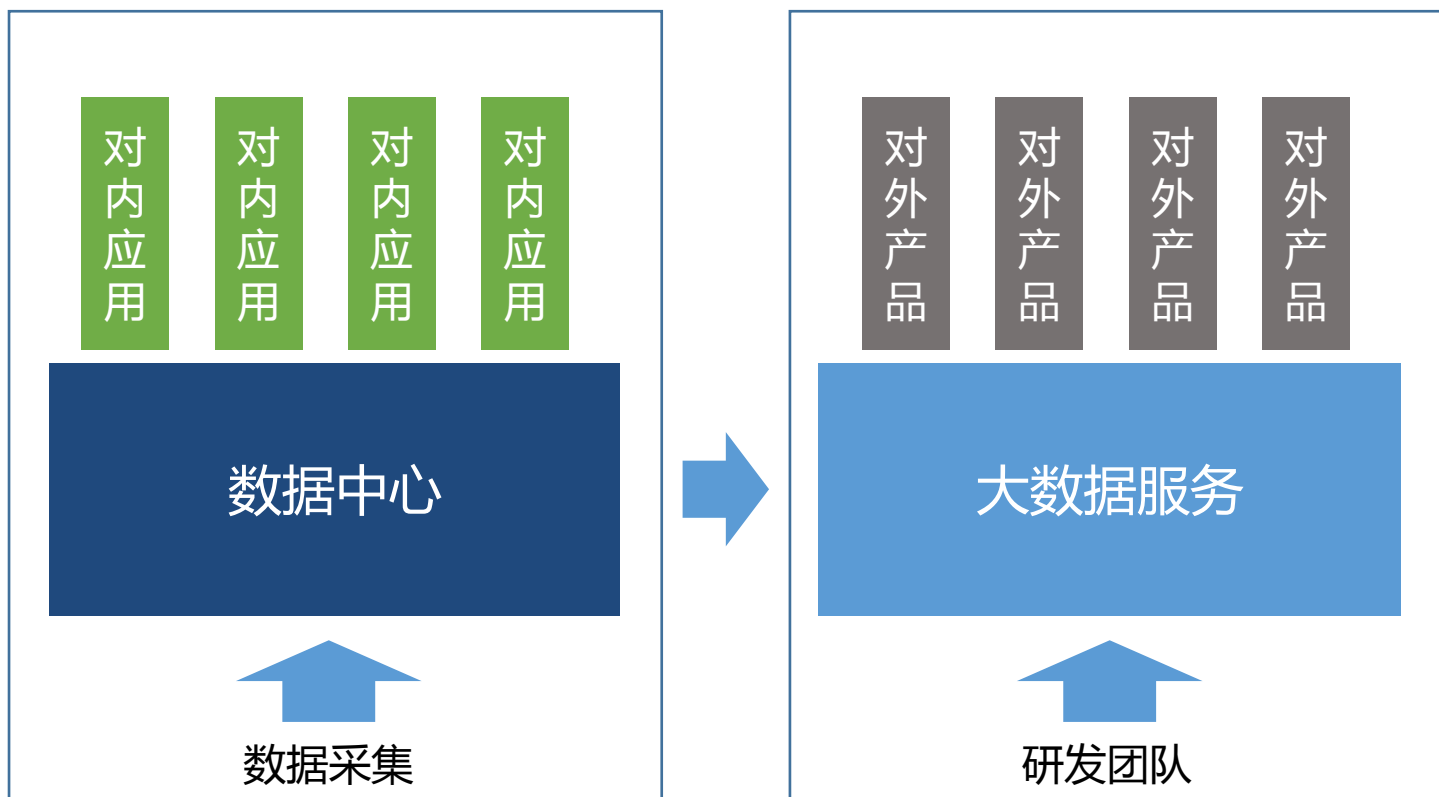


**2016年**

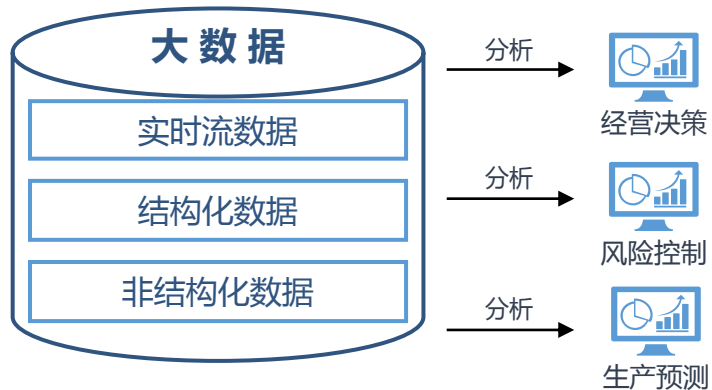


数据中心从成本中心转变为**利润中心**，向外部提供服务

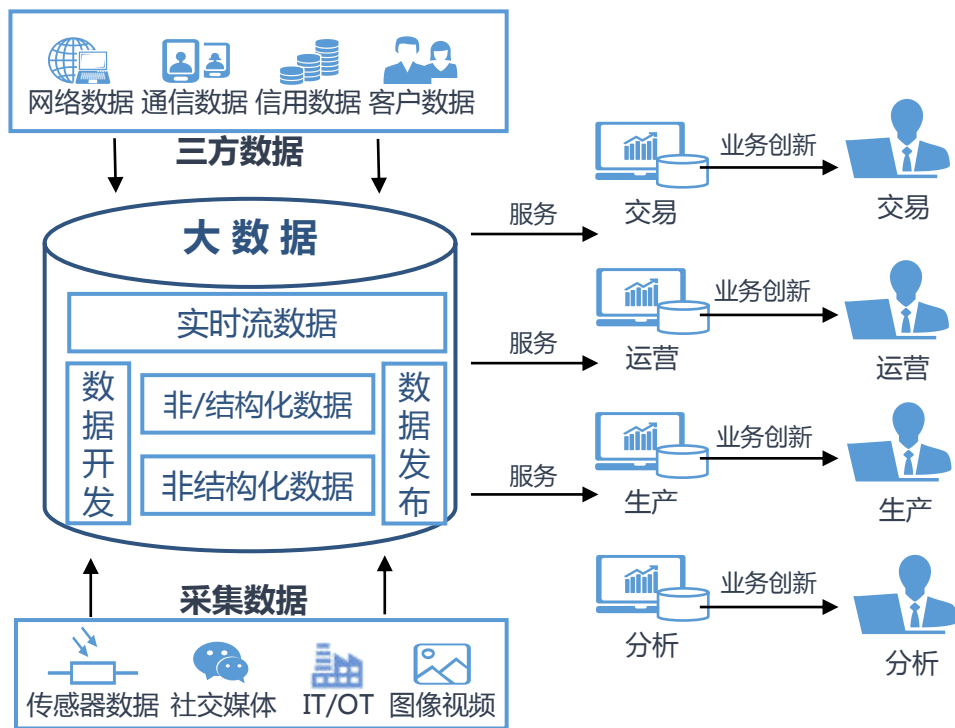
数据中心由数据管理转变为**数据服务**，为业务创新服务



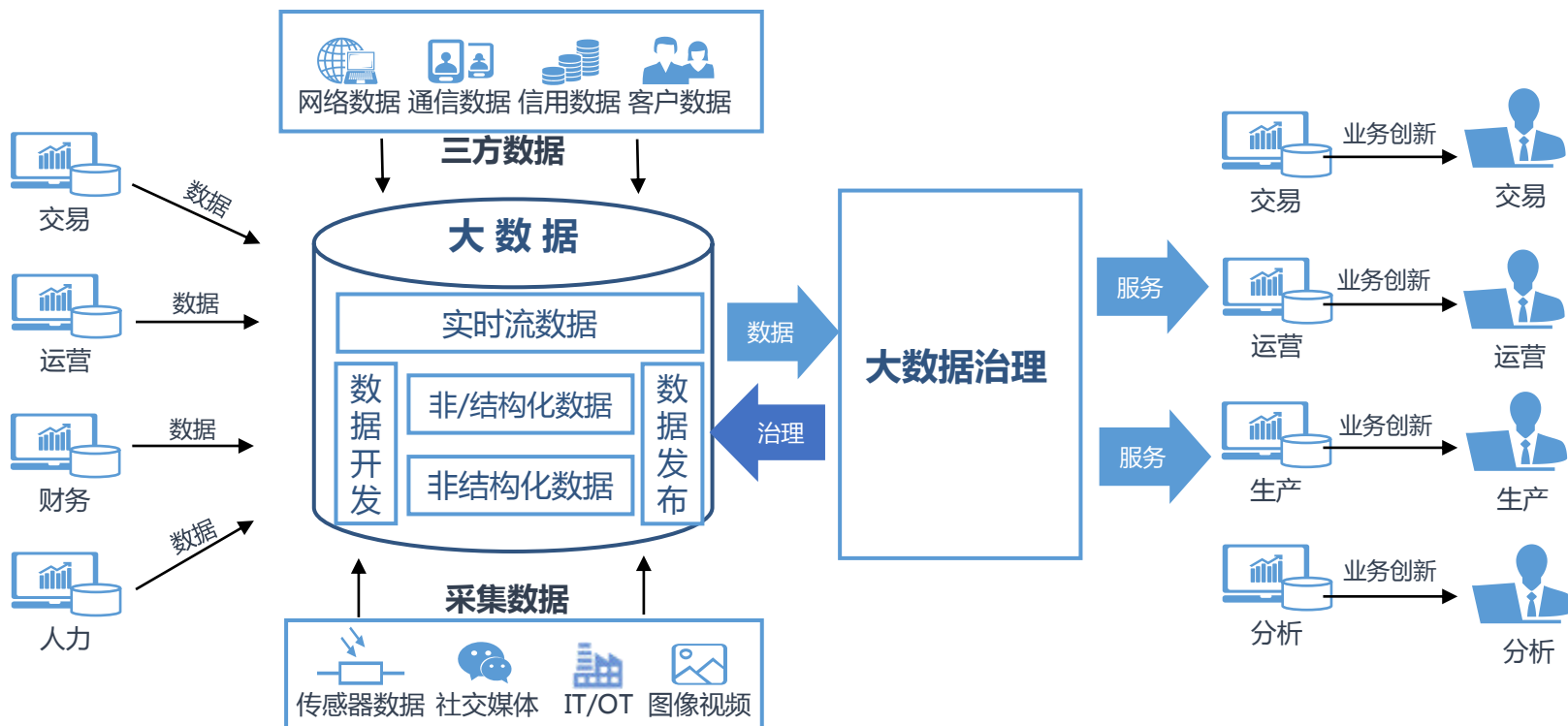
## 面向分析业务



## 为全业务域业务创新需求提供服务

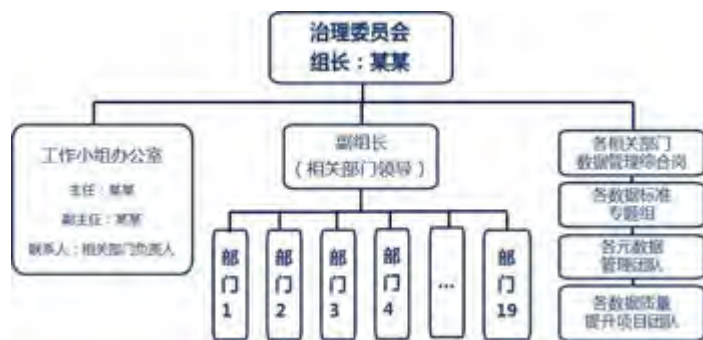


## 用大数据治理连接大数据与业务创新

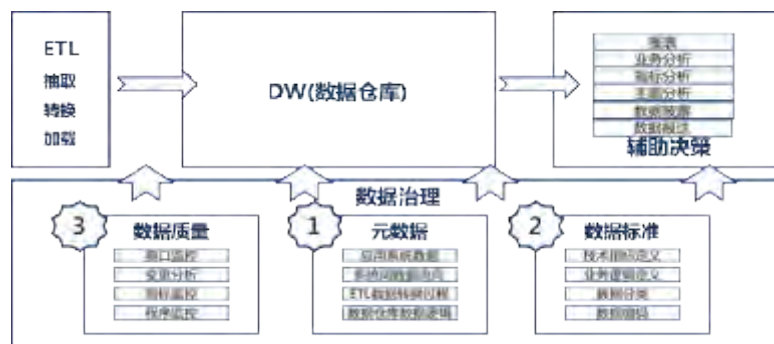


传统数据治理存在组织架构**复杂**、管理范围**狭窄**、管理流程**复杂**、人工录入多等弊端

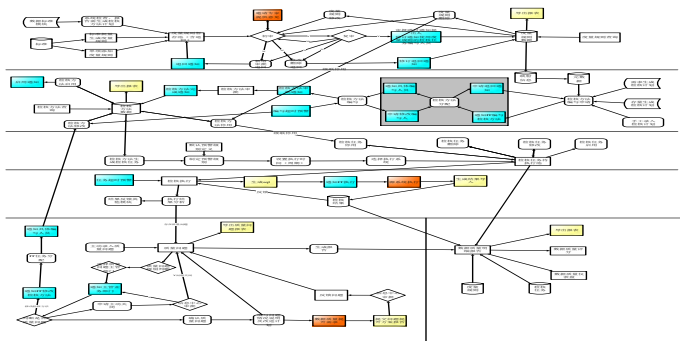
## 庞大的组织架构



## 狭窄的管理范围



## 复杂的管理流程



## 大量的录入工作

版本 1.1						
主题	信息项名称	英文名称	主管业务部门	描述	版本	标准编号
					1.1	

关联系统					
系统名	C 创	R 读	U 更		
系统 1					
系统 2					

待修改数据标准信息项初审					
包含标准项					
主题	标准名称	英文名	主管业务部门	标准描述	标准编号

表单详见 1.12.5.2

比对版本

通过初审

退回申请人

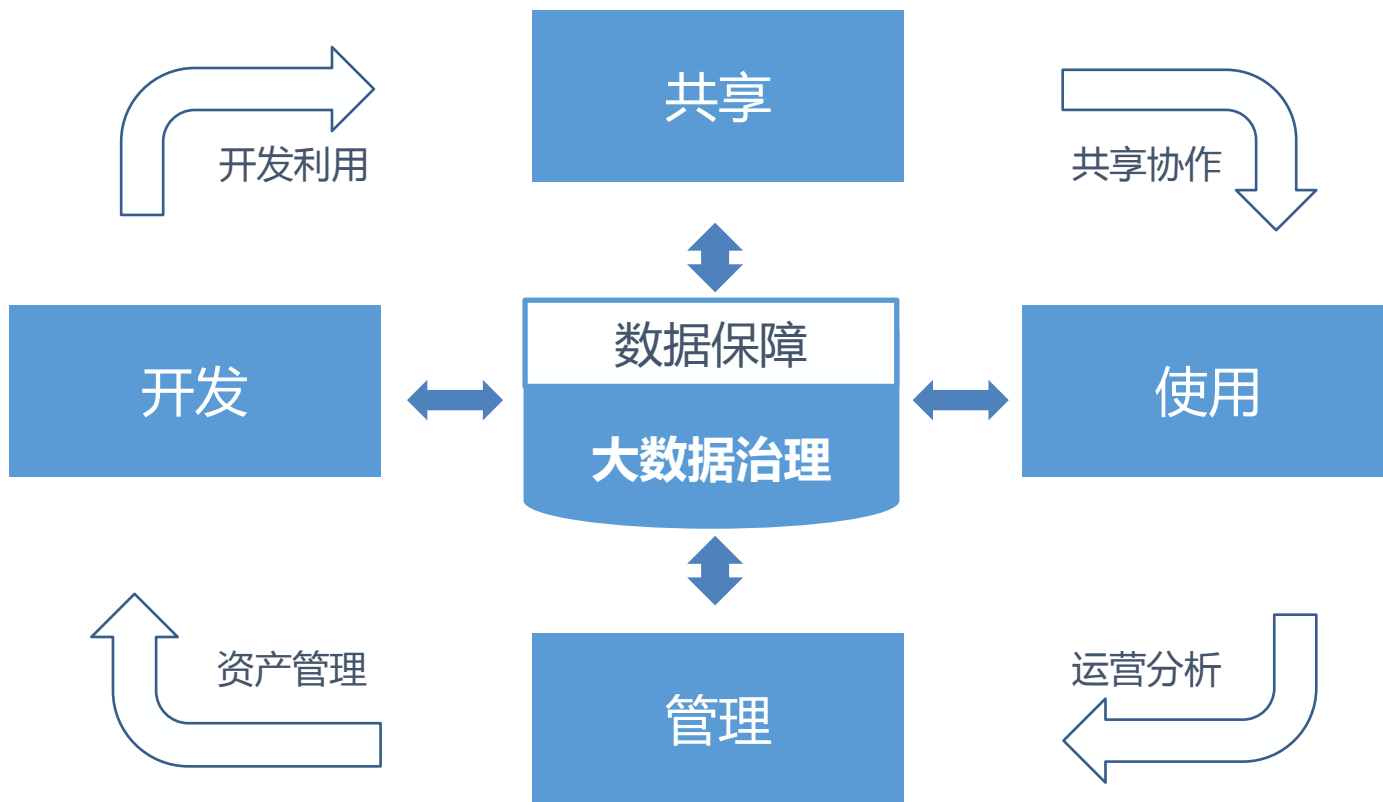
上传附件

表单详见下表	
确认	修改标准代码

用**自动**化，**自助**化，**智能**化的大数据治理

**找**到数据，**提**供数据，**使**用数据，**治**理数据



**1** 数字化时代大数据向服务化发展

**2** 大数据治理架构与关键技术分析

**3** 普元大数据治理架构实践





管理维度

缺乏服务化



业务维度

数据难使用

数据难查找



技术维度

数据工具链复杂

数据质量难提升

敏感数据易泄露

## 大数据治理



通过大数据治理，提供**数据服务**，从管理中心转变为**服务中心**

通过大数据治理，变现**数据价值**，从成本中心转变为**利润中心**

## 传统数据管理

管理  
职责



数据管理



数据开发



数据应用

管理  
思路



自己应用

管理  
组织



业务



开发



运维

全栈能力团队

## 创新数据管理



提供数据服务



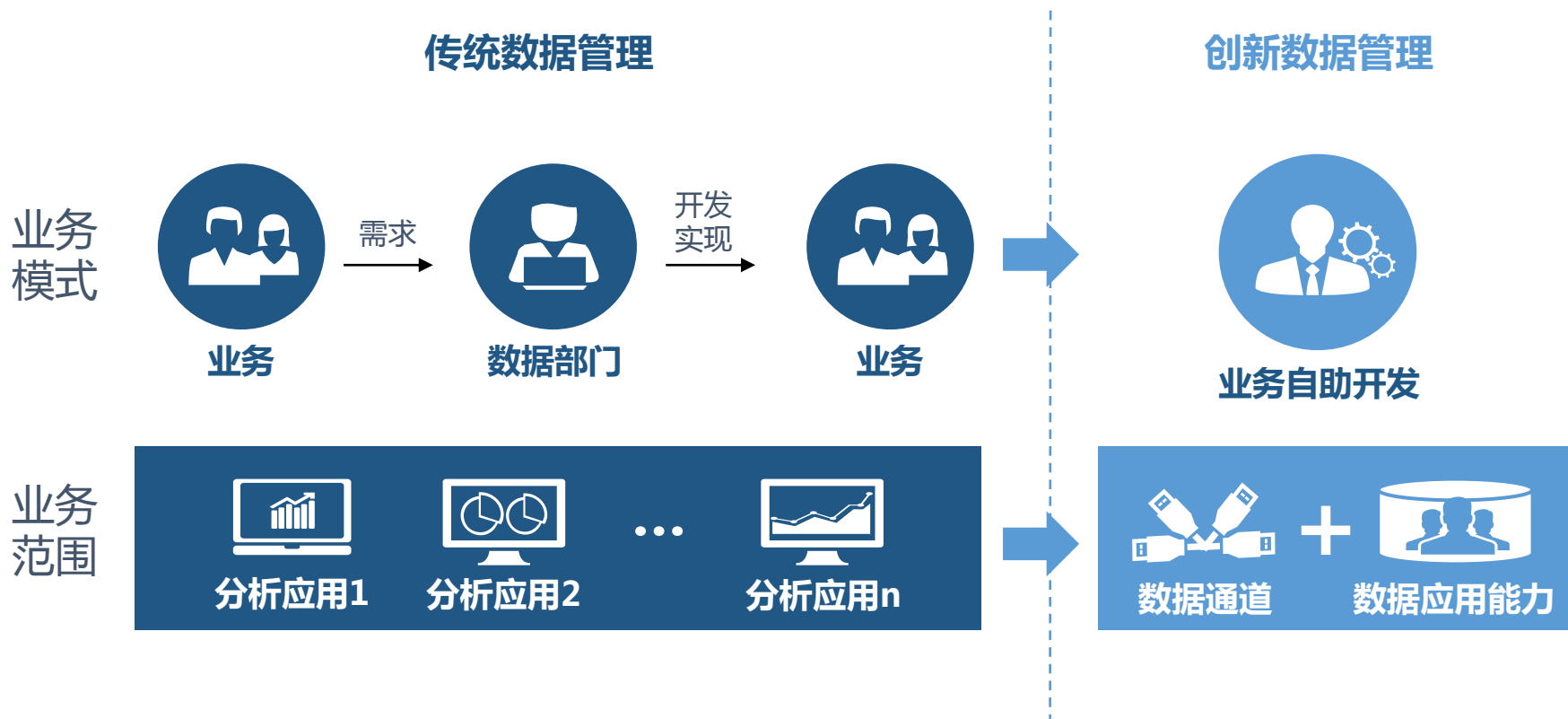
方便别人

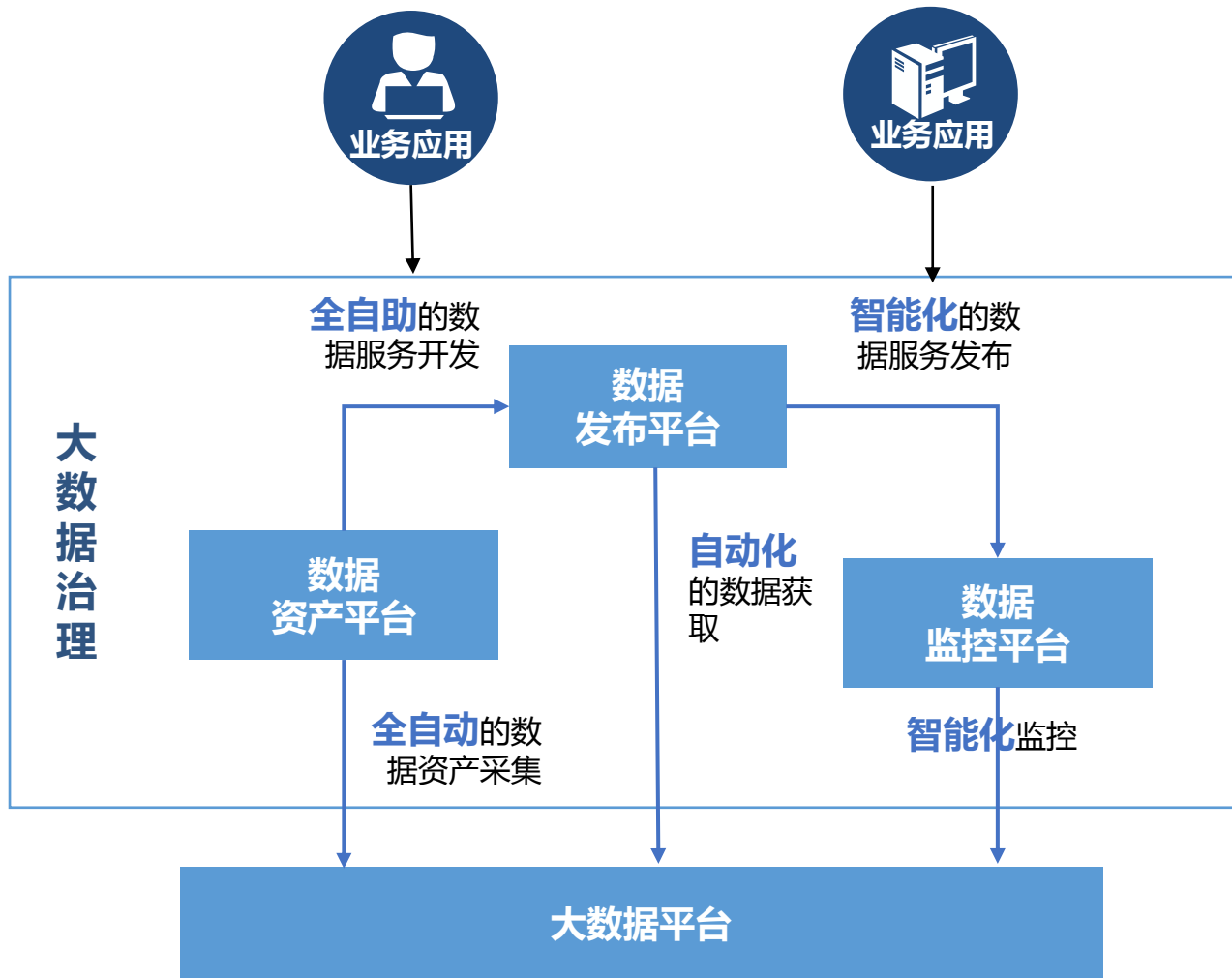


专业的服务团队



通过**大数据治理**，降低业务对技术依赖，充分发挥**业务创新**潜能





1

可视化管理企业资产

2

自动获取数据信息

3

大数据服务智能化

4

数据管理能力服务化

5

从需求开始控制数据质量

6

在集成点检查数据质量

7

持续积累检核规则

8

自动化质量评分

9

管理核心数据定义

10

为数据标准添加技术属性

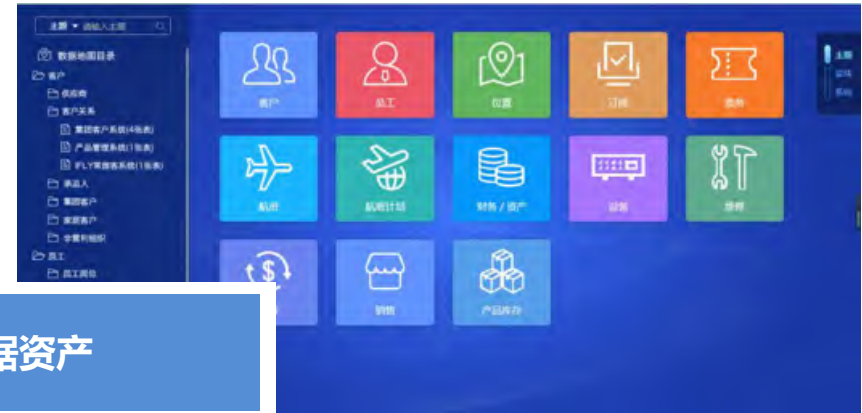
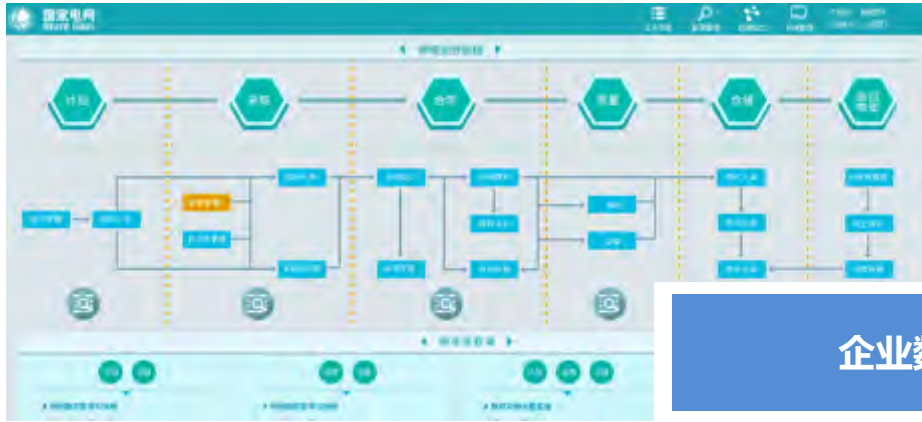
11

识别与管理业务语义

12

关联业务语义与技术



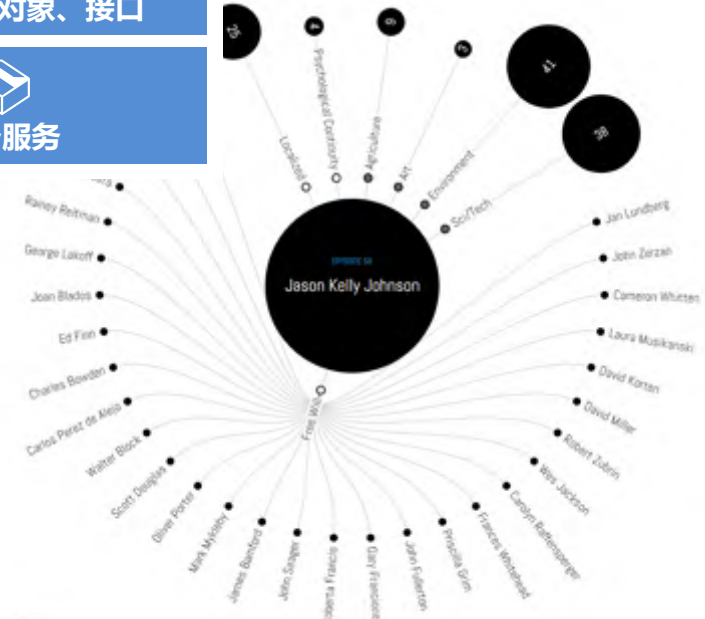
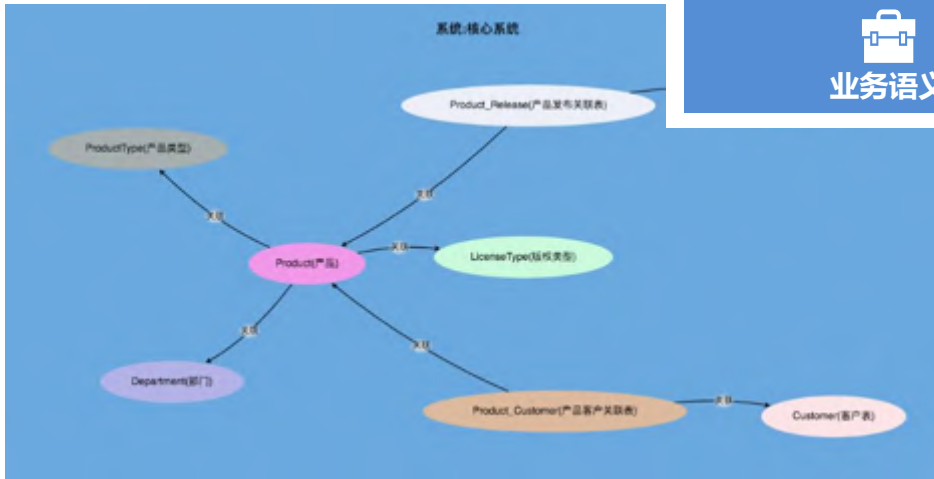


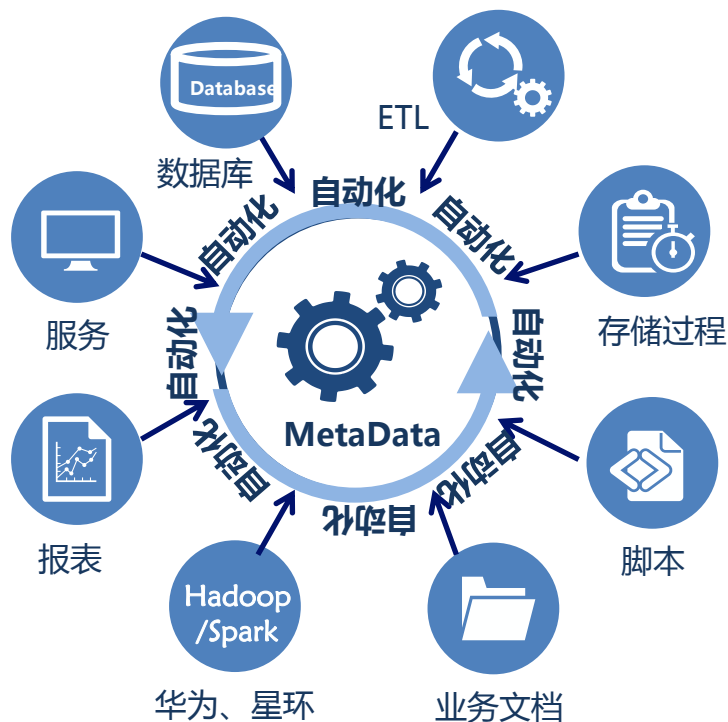
## 企业数据资产

表、字段

对象、接口

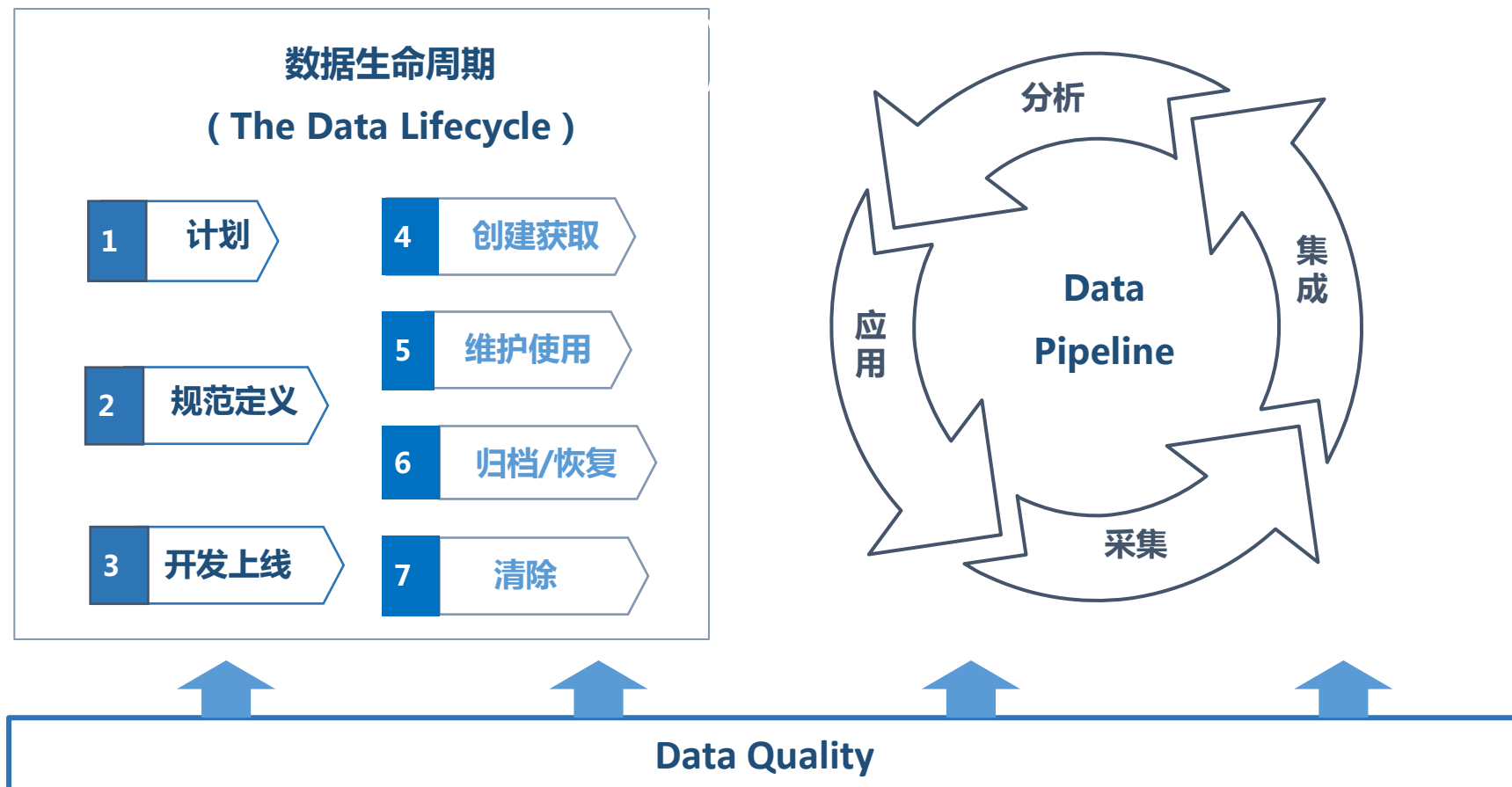
业务语义、业务服务





大数据适配器	Spark , HDFS , Hive , HBase , Sqoop , Flume , Pig
商业版本适配器	Transwarp Data Hub , Cloudera CDH , FusionInsight , Hortonworks HDP
数据库适配器	Oracle , DB2 , MySQL, Teradata , Sybase , SQLServer , Greenplum , PostgreSQL等多种数据库
建模工具适配器	ERWin , Powerdesigner等设计模型
ETL工具适配器	Primeton DI , Datastage , PowerCenter , Pentaho Kettle , SAP Data Serive
可视化工具适配器	Primeton iData , Cognos , OBIEE, SAP BO , Tableau , MSTR , QlikView
脚本适配器	对Shell、Perl等脚本采用SQL解析的方式进行ETL作业映射关系元数据采集, 存储过程, DDL
NLP适配器	自动分析各种非结构化文本, 自动分析业务元数据

以元数据为核心，打通数据生命周期，从源头保证数据质量。





提供各种数据管理服务，如质量、元数据、标准、监控、发布共享的服务，  
将治理平台变为服务平台。

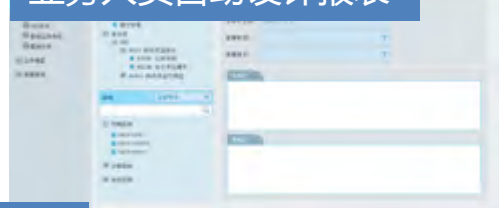
标准化服务帮助  
设计应用数据模  
型



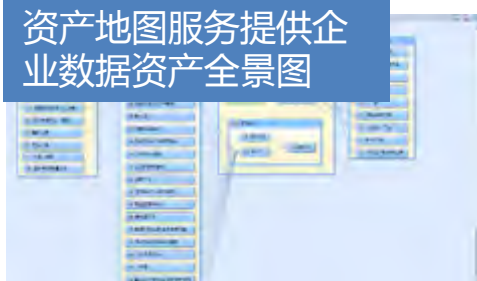
质量评分服务评  
估业务服务质量



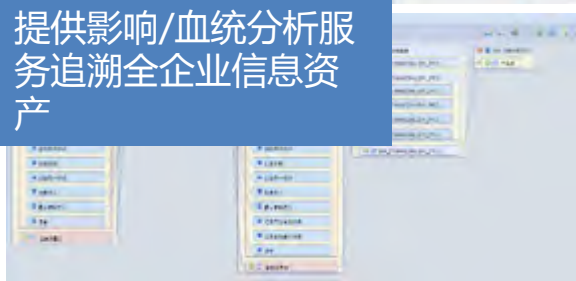
业务与技术关联服务使  
业务人员自助设计报表



资产地图服务提供企  
业数据资产全景图



提供影响/血统分析服  
务追溯全企业信息资  
产



## 能力服务化

元数据服务

数据质量服务

数据标准服务

数据开发服务

数据监控服务

数据发布共享服务

数据管理平台



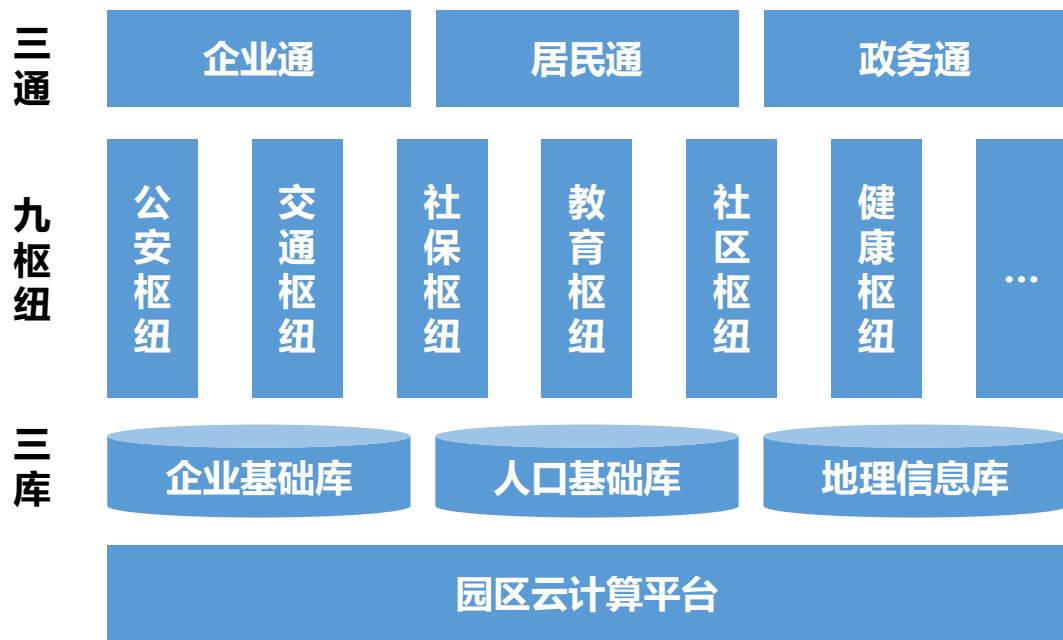
**1** 数字化时代大数据向服务化发展

**2** 大数据治理架构与关键技术分析

**3** 普元大数据治理架构实践



苏州工业园区在信息化发展规划中提出了“三库、三通、九枢纽”的信息化总体发展框架。

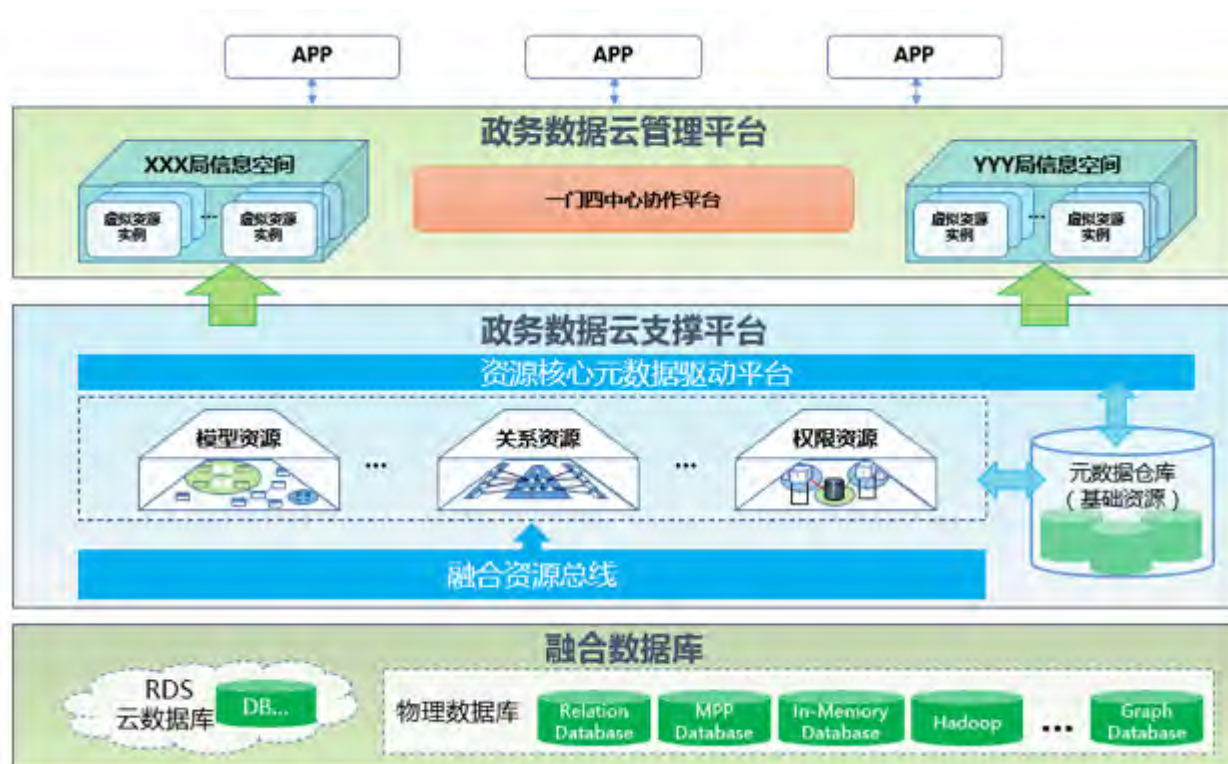


接入园区**31**个部门、涉及**85**个业务领域，**335**个数据分类，**3981**工作项。

先后完成了**公安**、**工商**等**15**家**人口**、**法人**信息高关联度局办的信息纳入政务资源云平台。

实现以元数据为基础的政务云**业务创新**模式（资产、开发、运营、监控）四中心  
一个技术支撑平台，一个融合库，全面实现苏州工业园的**数据资产资源化**

- ✓ 以元数据为基础梳理服务目录
- ✓ 将服务目录编目并与数据对接
- ✓ 将下属企/事业单位的数据能力接入政务资源云平台

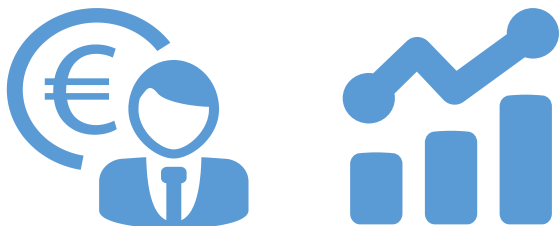


普元提供的数据自助分析平台，让阿里大数据平台内的**数据业务化**，成为数据存储平台与业务人员之间**沟通的桥梁**，让业务人员可**直接使用**大数据平台内的数据。

- ✓ 梳理业务元数据，形成业务元数据的统一管理
- ✓ 将业务元数据和技术元数据对应
- ✓ 自动形成业务化的数据地图
- ✓ 针对每个业务项对查询分类，帮助业务用户自助获取数据

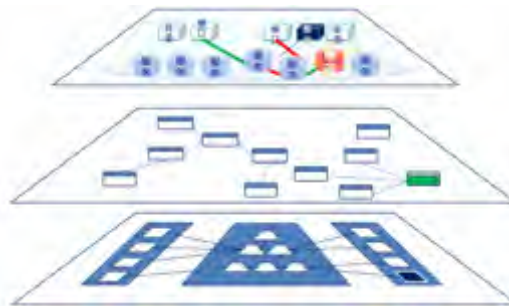


## 1.实现业务人员自助分析 提高效率



实现业务人员的自助分析，有业务人员直接从数据平台制作报表的比例由原来的15张扩大到**110**张，大大节省IT支撑人员使用成本。

## 2.建立全方位的业务应用视图 统一标准



通过余**2000**项业务数据项的梳理，建立数据视图，规范数据资产管理体系。规范化业务数据标准，推动国网数据字典梳理体系的**落地化进程**，统一浙江电网**35**个主要项数据标准。

基于业务元数据的**业务数据地图**，使业务方能迅速获取数据



**13**个主题域，即客户、员工、位置、订座、票务、航班、航班计划、财务、资产、设备、维修、运价、销售、产品库存。

针对每个主题域给出了业务实体目录及定义，实体总数**227**个。

通过对数据仓库贴源层业务系统表的核查及分析，给出了每个业务实体对应的数据库表与系统。





**1 , 企业数字化转型需要大数据服务化**

**2 , 大数据治理是大数据服务化的基础**

**3 , 大数据治理需要有完整的技术平台支撑**

**4 , 大数据治理的技术平台需要满足十二原则**



## Web/Mobile

- Web应用
- 移动应用
- React APP
- API应用

## BPM

- 流程梳理
- 流程编排
- 业务活动监控
- 通讯使能BPM
- 移动BPM
- 社交BPM

## Integration

- API网关
- 作业调度
- 服务总线
- 事件总线
- 数据总线

## Data

- Spark
- 复杂事件处理
- 数据展现
- Redis缓存
- Hadoop

## Data Lake

- 元数据
- 数据质量
- 数据标准
- 数据资产
- 数据发布
- 数据地图
- 数据开发

## DevOps

- 身份认证
- 项目管理
- 团队管理
- 计划管理
- 看板
- 沟通管理
- 知识库
- 产品管理
- 配置管理
- 版本控制
- 持续集成
- 持续发布
- 持续测试
- 持续改进

## 服务框架

- 服务配置
- 服务路由
- 服务发现
- 服务注册
- 服务容器
- 服务治理
- 服务日志
- 服务性能监控
- 负载均衡
- 消息中心
- 资源调度

## 基础设施

### 环境服务

- 虚拟机
- 网络
- 存储
- 容器
- 文件

## 云计算产品开放式研发

### 开放内容

研发过程  
向内外公开



技术组件  
开放研发



研发成果  
向社会公  
开



### 开放形式

微信群

技术文章

在线培训

线下沙龙

微课堂

技术视频

扫描二维码加群



普元信通

# 点亮业务创新时代

普元大数据治理平台

**BDTC** 2016中国大数据技术大会  
Big Data Technology Conference 2016