



大数据PaaS平台解决方案 (迪数产品)

深圳天源迪科信息技术股份有限公司

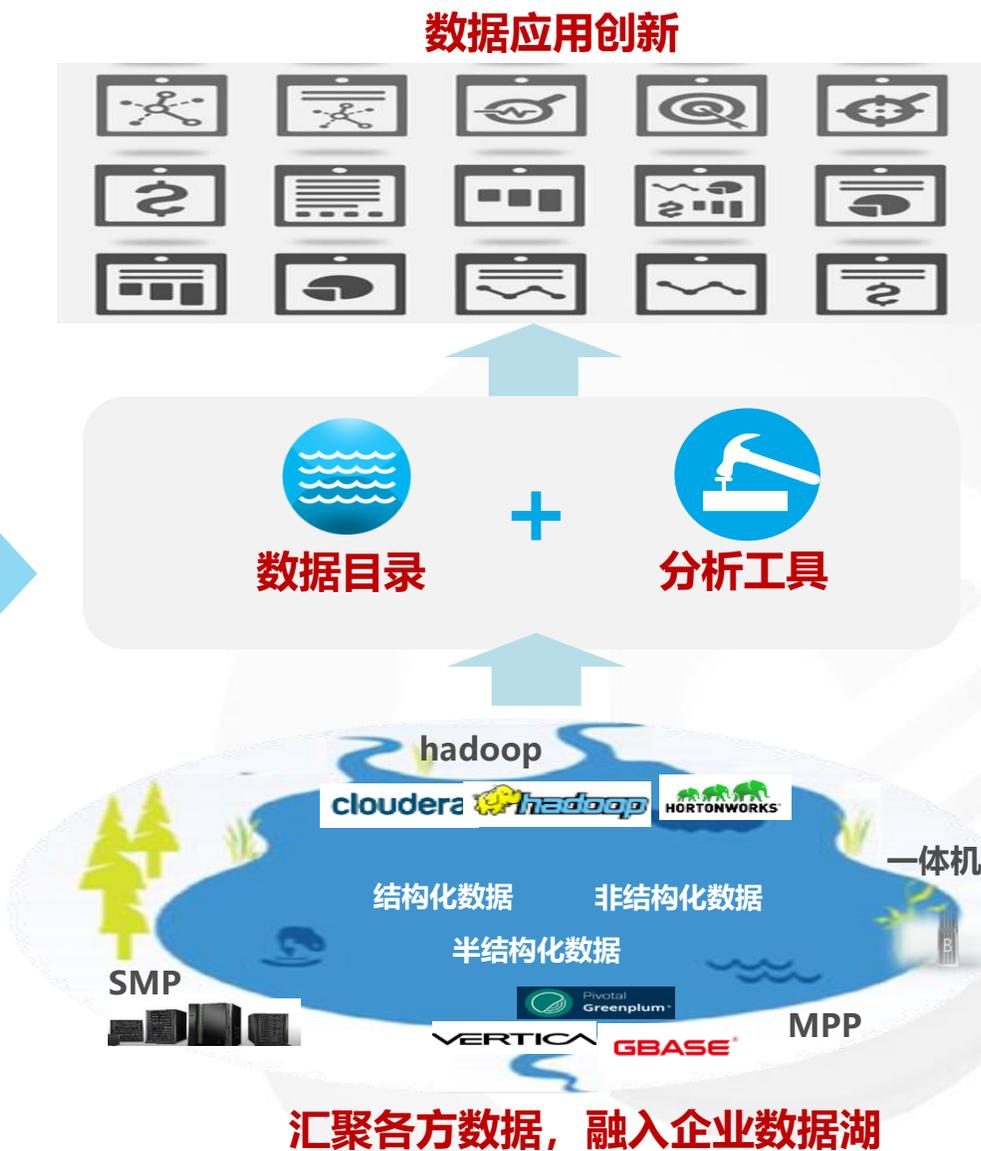
面向数据湖架构解决当前存在的问题

当前存在的问题

- 数据应用创新难
- 技术创新成本高
- 开发代码和知识无法沉淀
- 缺少统一数据分析平台
- 企业内部的信息存在壁垒
- 缺少统一数据标准
- 数据源类型结构繁多



希望不要求统一的格式，可供任意目的的分析



// 建设目标

大数据时代，企业急需建设企业级大数据平台，解决“**计算资源、存储资源、数据资源、工具及服务**”共享问题，**汇集各方数据，打造融合的数据湖，建设企业级数据目录，需要加强基础能力产品化和工具建设思路，实现企业数据和应用共享，降本增效**

打造面向未来、高性能、可扩展的，融合的数据湖



大数据平台具备“集中、开放、云化、产品化”要求



平台和应用分离，实现数据应用百花齐放



提供PaaS平台解决方案 (迪数)

数据应用

运营商行业应用

政府行业应用

金融行业

其他行业



多种部署模式

可以根据行业客户要求，提供
私有云和**公用云**两种部署模式



技术保持领先

技术保持领先性和独立性，系统架构
具有良好的**灵活性、可扩展性、安全性**

一站式体系化产品能力

提供一站式全套体系化产品能力，
覆盖了**商业智能、数据开发、数据
可视化**等领域

产品成熟度高

产品成熟度高，实现**集成化、
流程化、可视化和智能化**

天源迪科具备丰富数据处理经验

数据处理规模 (已商用全国级项目案例)

300+T

当前运行的单个案例
最大**非结构化**数据量

120+PB

当前运行的单个案例
最大**结构化**数据量

5000+个

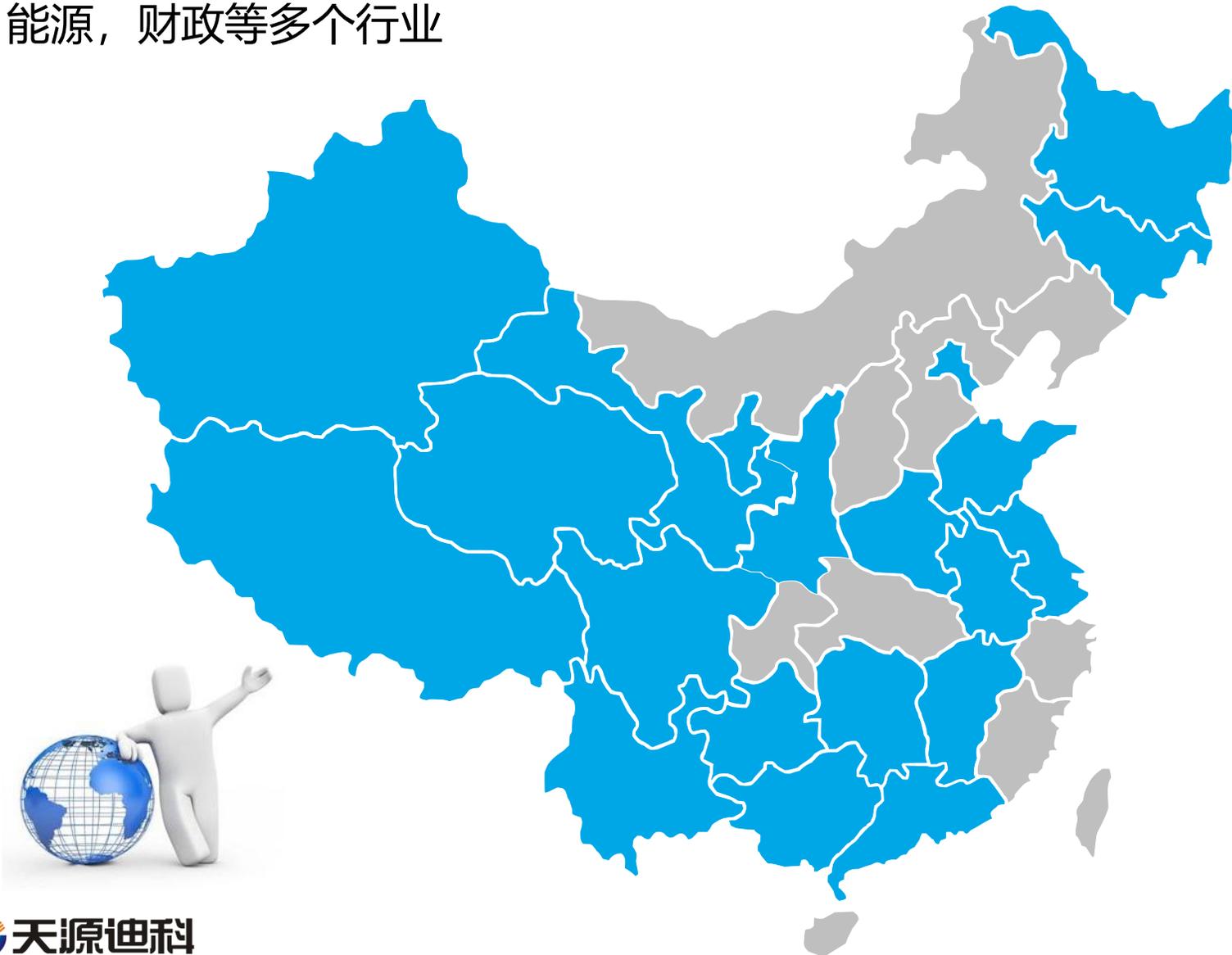
当前运行的单个案例
集群**数据处理节点**

数据处理技术种类

- ✓ 结构化:
- ✓ 半结构化: 上网行为数据、DPI、电商交易数据、电子邮件
- ✓ 非结构化: 视频、图片
- ✓ 流数据处理
- ✓ 离线数据处理

天源迪科大数据PaaS平台 — 客户分布

共**18**个省份，覆盖电信，移动，联通，公安，广电，新能源，财政等多个行业



项目使用

- 数据治理套件 **18+**
- 数据开发分析套件 **57+**
- 数据安全套件 **13+**
- 计算存储平台 **9+**

持续
增长中

大数据PaaS平台 — 为客户创造价值

为企业提供一站式大数据平台解决方案，帮助客户降低创新成本,助力企业在DT时代更敏捷、更智能、更具洞察力



提供一条龙数据开发和运营服务

研发全流程支持

多团队协作开发



大规模
计算处理能力

系统之间
数据融合

一站式的
数据工场

Web化的
软件服务

开放的
平台

大数据集群能力
轻松处理海量数据

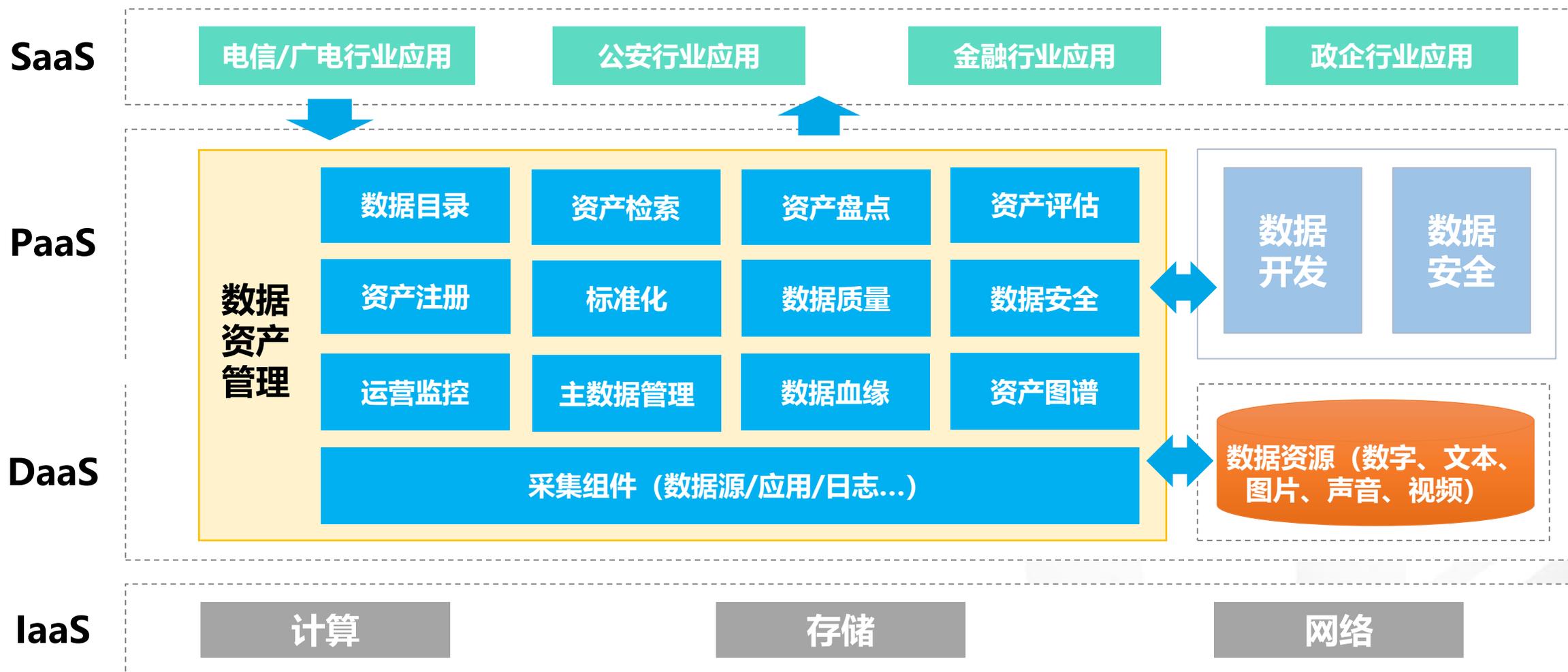
汇聚各方数据
打造融合数据湖

提供可视化工作
流程设计器功能

无需安装部署,
开箱即用

所有模块已
实现组件化、服务化

数据治理套件 — 数据资产解决方案



数据治理套件 — 数据资产盘点 (资产检索和血缘分析)

主数据管理

元数据管理

数据表管理

数据检索

数据目录

数据类目管理

数据资产视图

数据管理/数据检索/详情

数据检索-详情

模型详情 血缘分析 影响分析 **实体关联** 变更日志 产出信息 稽核规则

■ 实体表 ■ 维表 ■ 标签 ■ 指标 ■ 主表

```
graph LR; A[default.off] --> B[default.off]; A --> C[default.tb]; B --> D[default.prod_inst]; C --> E[default.c]; D --> E;
```

inst_id	实例标识
r_cust_id	客户标识
id	标识
create_date	时间
update_date	时间
delete_date	时间
order_id	订单号
valid_eff_date	生效时间
valid_exp_date	失效时间
product_type	产品类型

rel_id	关系标识
inst_id	实例ID
inst_type	实例类型

accounting_period	账期
inst_id	实例ID
item_type_id	类型
duration	时长

inst_id	实例ID
inst_name	实例名称
inst_type	实例类型
parent_mode_cc	父级模式标识
r_cust_id	客户标识
cust_id	实例标识

inst_id	实例ID
area_code	区域代码
partner_id	合作伙伴标识
duration	时长
charge	费用
inst_id	实例ID

开发分析套件 — 能力开放门户

能力开放门户 产品 应用市场 人工智能 解决方案 文档与支持 天源迪科

全部产品

- 大数据基础资源
- 分布式计算
- 分布式组件
- 大数据开发套件
- 数据分析及可视化
- 微服务网关
- 服务开发
- 服务治理
- 人工智能
- 数据治理
- 运维管理

大数据基础资源

- 存储资源
- 计算资源
- 多租户

分布式计算

- 离线计算
- 实时计算
- 消息队列

分布式组件

- 分布式缓存Redis
- 分布式消息系统Kafka
- 分析性数据库
- 分布式文件服务
- 图形数据库Neo4J
- 分布式搜索引擎
- 分布式爬虫

大数据开发套件

- 数据集成
- 数据上传下载
- 批处理开发
- 实时开发
- 智能调度
- 服务封装

数据分析及可视化

- 即席查询
- 实时多维分析
- 文本挖掘
- 报表工具
- 可视化大屏

微服务网关

- API网关
- 负载均衡
- 消息服务

服务开发

- 服务封装
- 服务注册
- 服务整合
- 服务发布
- 服务下架

服务治理

- 服务接入
- 服务鉴权
- 服务优先级
- 服务熔断
- 服务计量

人工智能

- 自然语言处理
- 图像识别
- 文字识别
- 人脸核身
- 人脸识别
- 聊天机器人

数据挖掘

- 机器学习
- 深度学习
- 特征工程
- 统计分析
- 自动建模

数据治理

- 数据注册
- 数据标准化
- 数据质量
- 数据血缘
- 资产地图

运维管理

- 监控中心
- 安全审计
- 日志分析

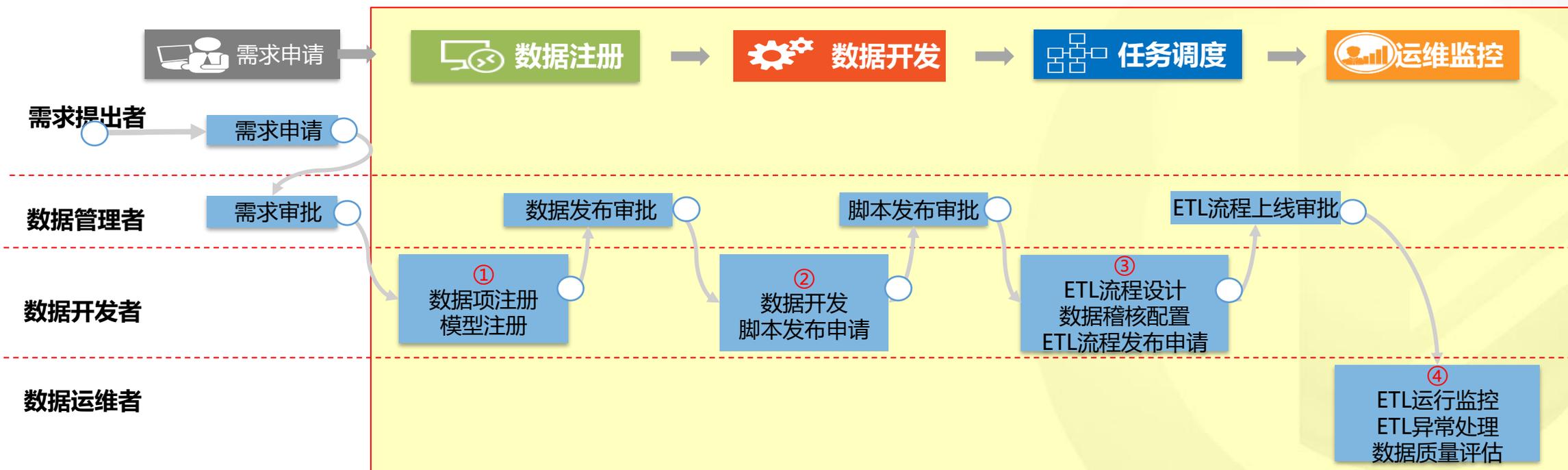
开发分析套件 — 产品价值及定位

产品定位 为企业大数据实施提供的一套完整的一站式大数据解决方案

产品价值 从数据导入、查找、开发、ETL、调度、部署、建模一站式集成开发环境，降低数据创新成本

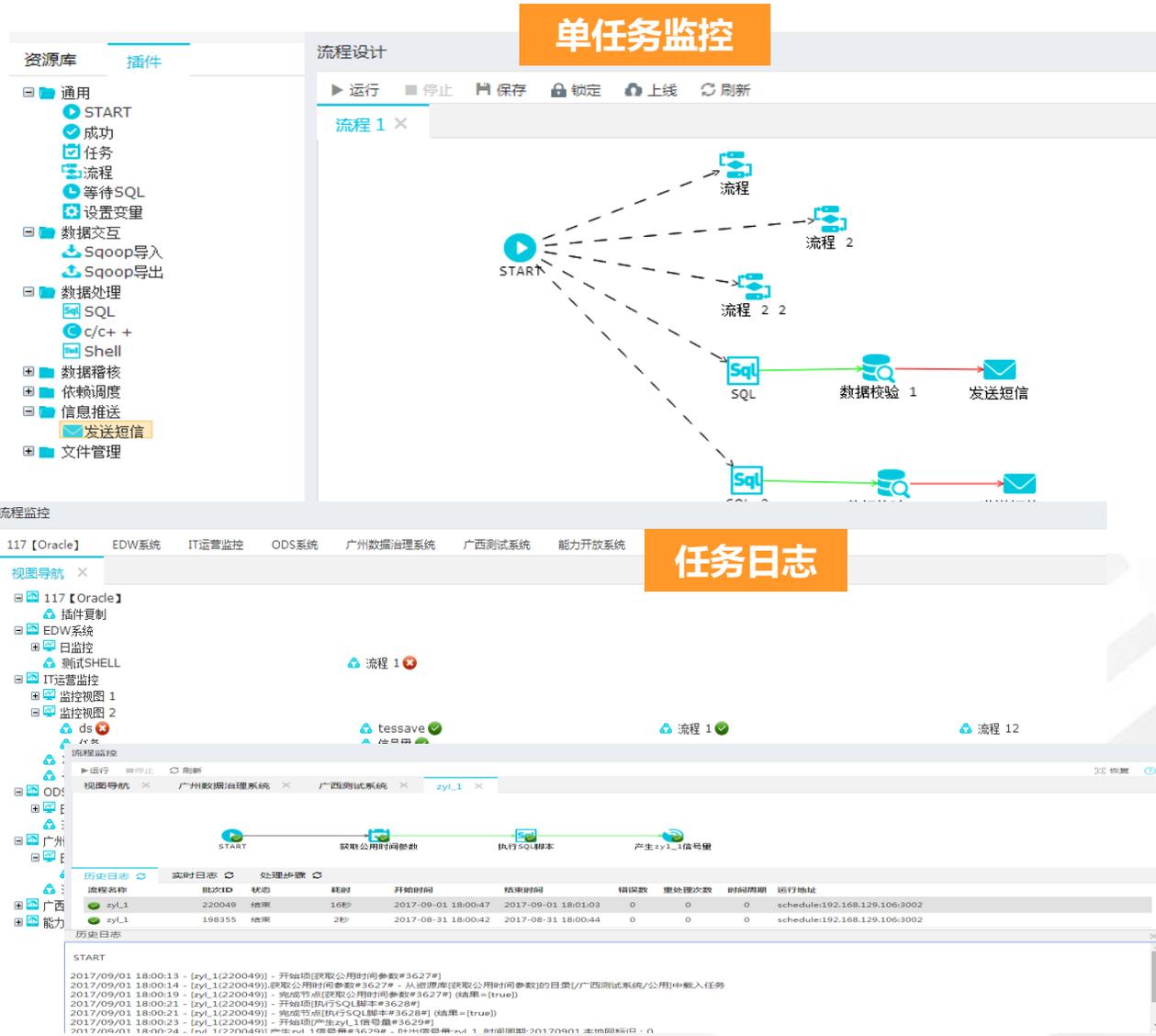
数据管理、开发和运营一体化

多
用
户
场
景

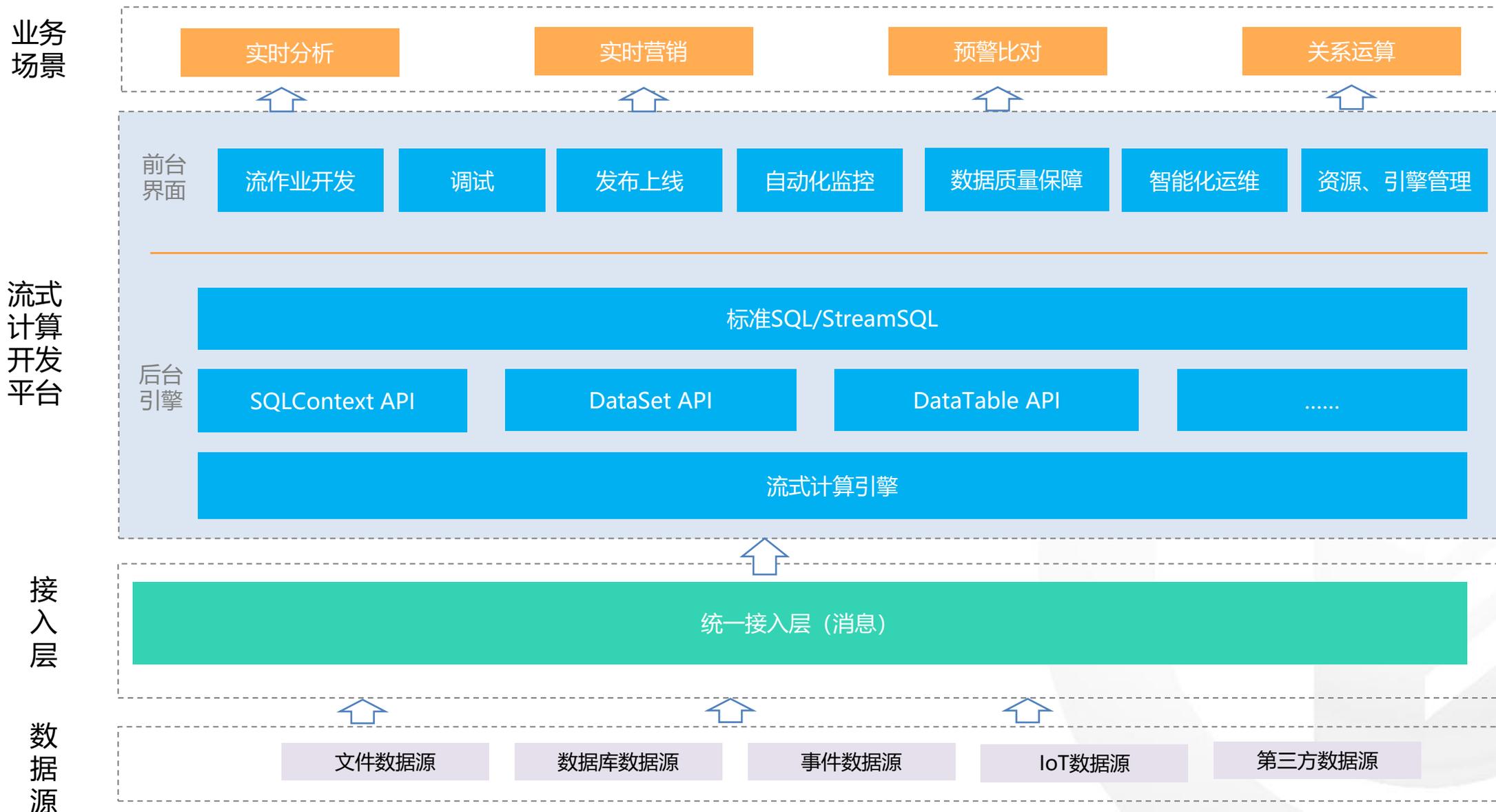


开发分析套件 — 任务调度和监控

提供分布式调度引擎、监控视图、数据稽核报告、异常推送和异常重处理等能力，为企业各类批量处理任务执行提供支撑，与数据开发工厂贯通，提高工作效率



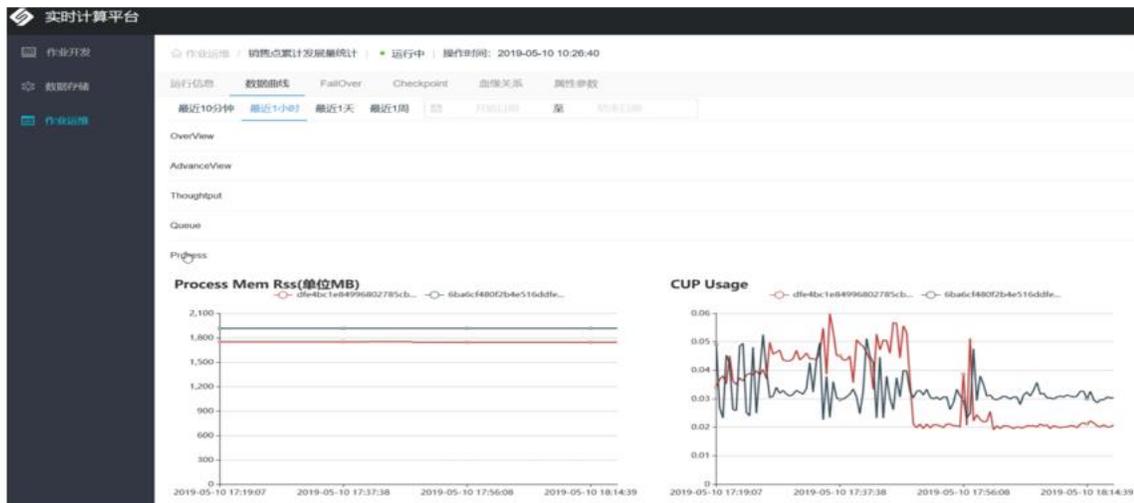
计算存储套件 — 实时计算平台



计算存储套件 — 构建一站式在线流式开发

面向高吞吐、低延时、准确性和易用性场景的一站式实时开发工具。引入Flink计算框架，支持多租户管理、SQL化开发、在线调试、一键发布、热生效和自动化监控运维，满足企业实时分析、实时营销、实时预警、实时比对等数据应用场景

运行监控



脚本开发

```
64 prod_date date,  
65 operators_id long  
66 ) WITH(  
67   'dataSource' = 'mysql://155729860202e35237ac434478598e223d9ce1396ed',  
68   'primaryKeys' = 'channel_id'  
69 );  
70  
71 --临时表 (用 prod_inst 表关联)  
72 CREATE TEMP STREAM prod_inst_1(  
73   prod_inst_id varchar,  
74   prod_inst_name varchar,  
75   ser_nbr varchar,  
76   status_cd varchar,  
77   staff_id varchar,  
78   acct_id varchar,  
79   channel_id varchar  
80 );  
81  
82 CREATE RESULT TABLE prod_chan_1(  
83   prod_inst_id varchar,  
84   prod_inst_name varchar,  
85   ser_nbr varchar,  
86   status_cd varchar,  
87   acct_id varchar,  
88   staff_id varchar,  
89   channel_id varchar,  
90   channel_name varchar  
91 ) WITH(  
92   'dataSource' = 'mysql://155729862290161c9c4991e12408818953e7a1b2871b',  
93   'primaryKeys' = 'prod_inst_id',  
94   'batch.size' = '100',  
95   'batch.interval.second' = '3'  
96 );  
97  
98  
99  
100 INSERT INTO prod_chan_1 JOIN_S_1 SELECT  
101   a.prod_inst_id,a.prod_inst_name,a.ser_nbr,a.status_cd,a.staff_id,a.acct_id,b.  
102   from prod_inst a join tb_staff for system_time as of proctime() as b on a.sta  
103   ('100000','130000');
```

定义数据源

名称	存储类型	注册时间	创建人	修改人
tb_staff	mysql	2019-05-08 14:25:38	系统管理员	系统管理员

STAFF_ID	STAFF_NAME	PURTY_ID	STAFF_NAME	STATUS_CD	CREATE_DATE	SALES_CODE	LATH_ID	CRT_DA	EFF_DA	EXP_DA	MOGL_DA
100111	20	253278	李*强	1000	2015-1...	2014-0...	W9121...	560	2014-0...	2015-1...	2015-1...
100247	20	292061	李*强	1000	2013-0...	2013-0...	940371	560	2013-0...	2013-0...	2014-0...
100268	20	150729	赵*翔	1000	2013-0...	2013-0...	950056	551	2013-0...	2013-0...	2014-0...
100334	10	209632	罗*华	1000	2012-1...	2013-0...	AHAQ...	556	2012-1...	2013-0...	2013-0...
100375	10	300177	[*]	1000	2011-0...	2011-0...	AHRA...	564	2011-0...	2013-0...	2013-0...
100566	20	151187	徐*进	1000	2013-0...	2013-0...	950456	551	2013-0...	2013-0...	2014-0...
100591	20	274524	张*文	1000	2016-0...	2013-0...	960036	561	2013-0...	2015-0...	2016-0...
100766	20	1248060	蔡*红	1000	2013-0...	2013-0...	81259	560	2013-0...	2013-0...	2014-0...
100770	20	311862	王*强	1000	2013-0...	2013-0...	81351	563	2013-0...	2013-0...	2013-0...
101002	10	144573	[*]	1000	2013-0...	2013-0...	A140E	551	2013-0...	2013-0...	2013-0...

调试运行

谢谢聆听

T H A N K Y O U

更多信息，敬请关注微信公众号“天源迪科大数据”

无限数据，有形价值

天源迪科，帮您挖掘数据价值的专业伙伴！



深圳市高新技术产业园区南区南七道T3大厦B三层

ADD: 3/F,T3 Building,Nanqi Road,South Area Shenzhen Hi-Tech Industrial Park,Shenzhen, P.R.C

电话(TEL): +86-755-26745688

传真(FAX): +86-755-26745666

邮编(P.C.): 518057