



DataSimba

奇点云大数据智能服务平台

数智化转型的基础设施

www.startdt.com

2021-V1.0

目录

CONTENTS

1. 产品概述
2. 产品功能
3. 产品优势
4. 产品定位
5. 场景案例

01 产品概述

背景：企业消费者运营各环节均面临挑战



数据治理

针对各种来源的数据，**统一数据标准**、沉淀数据资产，同时保障数据**安全**，并促进数据在组织内**无障碍地应用**。



数据计算

汇聚处理各类业务数据，基于数据中台强大的计算引擎能力，完成**复杂的分析与监控工作**，**灵活响应并满足新需求**，实现对用户千人千面的精细化运营。



跨云多端

产业链的复杂性、业务模式和渠道的日益多元和碎片化，带来了**应用层显著的多端趋势和数据基础设施不可避免的跨云需求**。

数据中台需要解决「在跨云与多端情景下，统一的数据资产管理、数据价值生产和数据流动」的问题。

奇点云自研的一站式**大数据智能服务平台**，旨在为企业提供全链路的 **产品+技术+方法论** 服务。助力企业快速搭建安全、易用的数据中台，最大化释放价值，驱动业务增长与创新。

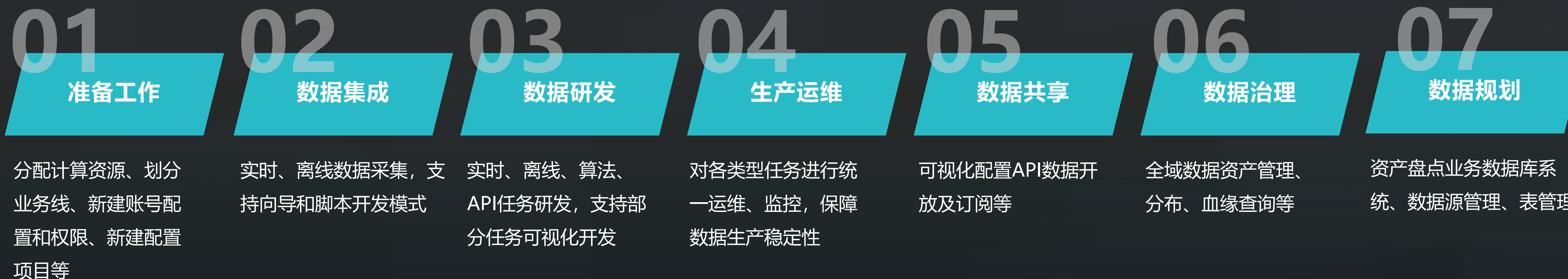
平台架构



02 产品功能

作为一站式大数据智能服务平台，为大数据开发的7个不同阶段提供一系列技术能力。致力于让企业高效挖掘数据价值，驱动业务增长与创新。

● 7大阶段



● 核心功能



1. 准备工作-分级多域数据治理

工作空间准备工作包含平台运维配置计算引擎、资源组、业务板块定义、创建项目、项目环境配置、接入数据源、子账号、角色。

配置计算和存储资源时，DataSimba支持分级多域数据治理技术，可以支撑同一个企业主体，配置多个不同的工作空间，工作空间之间的各IaaS厂商可进行隔离部署。

The screenshot displays the 'DataSimba | 工作空间控制台' interface. At the top, it shows the workspace name '生产域01区' and its status '已激活'. Below this, it lists the calculation engine (默认离线计算引擎、代理计算引擎、新测试计算引擎) and the resource group. The workspace creation time is 2020-12-09 17:10:44. A '业务规划 >' link is visible.

The main area shows a list of projects under the '离线项目' filter. Each project card includes:

- Project Name: proj_demonstrate, mz_ym, [redacted]
- Status: 离线 (Offline)
- Submitted/Total Tasks: 123 / 196, 0 / 0, 27 / 37
- Table Count: 196, 0, 25
- Storage Used: 2153.85MB, 0.00MB, 2575.89MB
- Today's Task Failure Count: 18, 0, 0
- Project Manager: demonstrate
- Creation Time: 2020-06-19 15:17:07, 2021-02-02 18:03:30, 2021-01-05 14:29:27
- Business Block: bs_demonstrate, 订单, 订单

Each project card has a '进入数据工作台 >' button. The interface also includes a search bar and filters for '创建时间' and '全部业务板块'.

2. 数据集成-多源异构数据集成

稳定高效的数据同步平台，支持20+种异构数据源通过增量或全量的方式，在各类复杂网络环境下进行高速稳定的批、流数据同步。

The screenshot displays the DataSimba web interface. At the top, there is a navigation bar with 'DataSimba' logo and menu items: '首页', '集成', '研发', '运维', '服务', '治理', '规划'. On the right, there are user and project information: '阿里云-生...', '离线 / proj_d...', and a profile icon.

The main content area is titled '欢迎使用离线数据集成' (Welcome to Offline Data Integration) with a subtitle '进行离线任务的数据导入，为后续的研发做数据准备工作' (Import data for offline tasks, preparing data for subsequent R&D work). Below this, there are two main sections:

- 支持 16 种数据源类型** (Supports 16 data source types): A grid of 16 data source icons including DataSimba, ORACLE, SYBASE, FTP, MySQL, SFTP, OSS, PostgreSQL, mongoDB, HBASE, HDFS, Impala, Tidb, Microsoft SQL Server, and HIVE. A blue button '+ 添加数据源' (Add data source) and a grey button '查看已添加的数据源 >' (View added data sources) are located below the grid.
- 支持数据文件本地上传** (Supports local upload of data files): A dashed box containing an icon of a folder with an upward arrow and the text '点击上传本地数据文件' (Click to upload local data files). A grey button '查看已上传的数据文件 >' (View uploaded data files) is located below.

At the bottom, there is a **离线数据集成使用流程** (Offline Data Integration Usage Process) diagram with four steps:

- 01 数据管理** (Data Management): 进行添加数据源或者数据文件 (Perform adding data sources or data files).
- 02 规范建表** (Standard Table Creation): 进行ods层原始表的创建 (Perform creation of original tables in the ods layer).
- 03 数据同步** (Data Synchronization): 进行同步任务的创建 (Perform creation of synchronization tasks).
- 04 数据运维** (Data Maintenance): 进行任务、实例监控 (Perform task and instance monitoring).

3. 数据研发-多类型任务构建

覆盖数据开发全链路的一站式数据中台解决方案，集成批/流数据开发能力，规范建模、标签工厂降低使用门槛，轻松完成多种类型任务构建，提升全链路研发体验。

规范建表流程说明

- + 建立业务过程: 代表一个被管理实体或系统的事实，是对业务活动事件的定义
- + 建立原始表: 将系统中的数据几乎无处理的放到数据平台中
- + 建立维度表: 维度是由独立不重叠的数据元素组成的数据集，所构成的可进行统计的对象
- + 建立明细表: 描述某个事实（业务过程）属性的数据仓库模型
- + 建立汇总表: 描述维度统计信息（即汇总指标）及维度属性信息的数据集，所构成的数据仓库模型
- + 建立应用表: 存放数据产品个性化的统计指标数据，面向需求的数据指标

表创建成功之后，会创建对应的空任务；ods层将在DATA_IMP中创建空的同步任务，其它层级将在DATA_TRSF中创建空的ETL任务

[前往参数配置 >](#)

4. 生产运维-智能监控管理

任务智能管理平台的所有发布任务分为周期任务和手动任务，平台可以对各类任务和对应实例进行统一运维、统一监控、统一管理，其目的是保障任务上线后的稳定、高效。



5. 数据服务-对外数据共享

数据服务是把数据变为一种服务能力，通过数据服务让数据参与到业务，进而激活整个数据中台，他包含可对外系统调用的API管理，可以看到某项目发布的API数，被调用次数等。

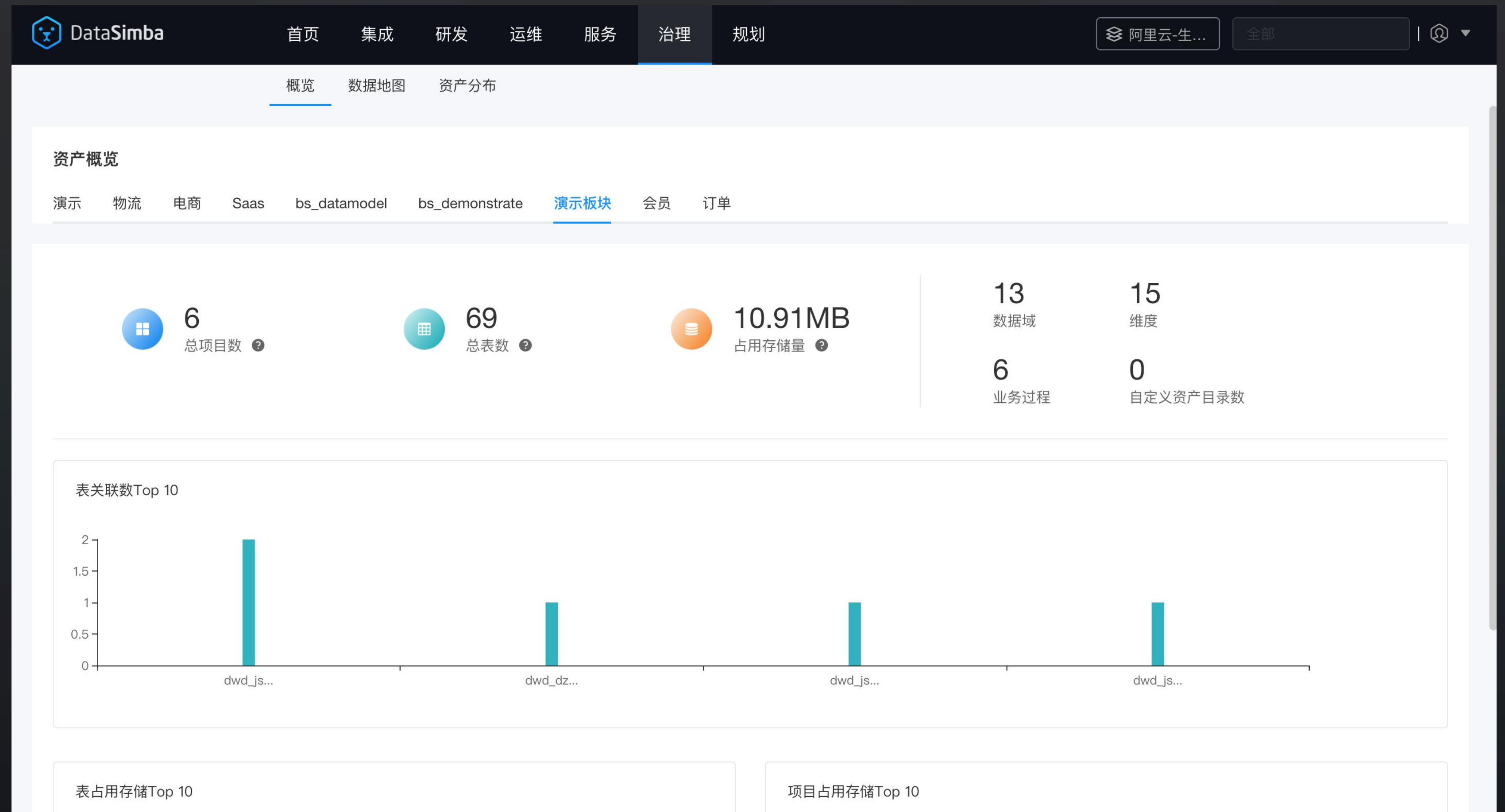
项目名称	项目状态	创建API数	发布API数	最近发布时间	累计调用次数	24小时内调用次数	24小时内
[模糊]	● 启用中	1	-	-	-	-	-
[模糊]	● 禁用中	4	-	-	-	-	-
11	● 启用中	6	1	2020-07-22 14:50:12	-	-	-
test	● 启用中	3	1	2020-07-23 18:21:32	-	-	-
识场	● 启用中	59	58	2020-03-10 10:54:46	-	-	-

共 11 条 5条/页 < 1 2 3 > 前往 1 页

6. 数据治理-元数据管理

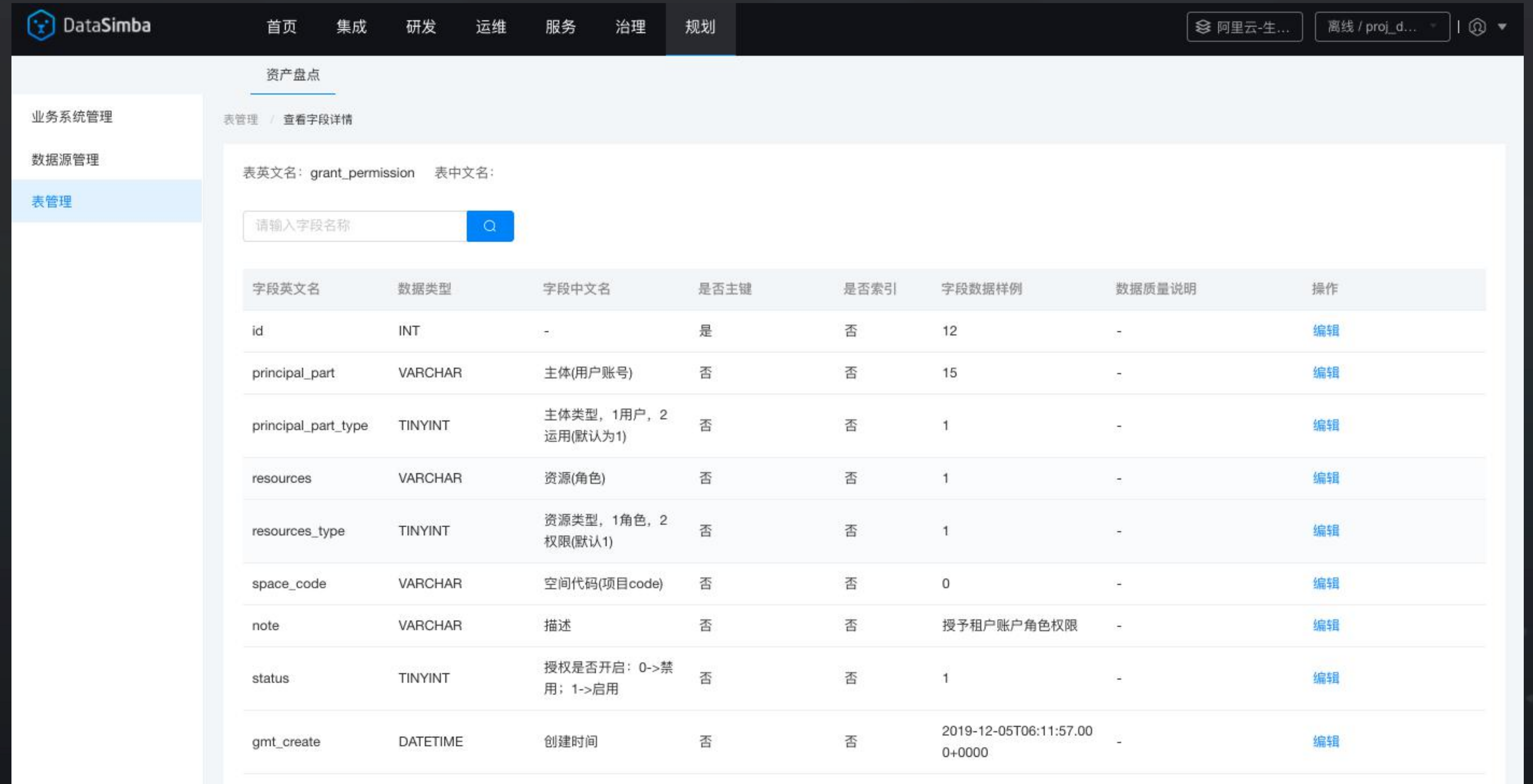
数据治理模块包含元数据管理规划和基于元数据的数据资产管理应用。

*元数据是描述数据的数据，按用途不同分为技术元数据、业务元数据和管理元数据。



7. 数据规划-业务系统盘点

数据规划支持对业务系统进行数据库、表、字段勘查，并进行周期同步和手动同步。



The screenshot shows the 'Asset Inventory' (资产盘点) interface in DataSimba. The table displays the following fields:

字段英文名	数据类型	字段中文名	是否主键	是否索引	字段数据样例	数据质量说明	操作
id	INT	-	是	否	12	-	编辑
principal_part	VARCHAR	主体(用户账号)	否	否	15	-	编辑
principal_part_type	TINYINT	主体类型, 1用户, 2运用(默认为1)	否	否	1	-	编辑
resources	VARCHAR	资源(角色)	否	否	1	-	编辑
resources_type	TINYINT	资源类型, 1角色, 2权限(默认1)	否	否	1	-	编辑
space_code	VARCHAR	空间代码(项目code)	否	否	0	-	编辑
note	VARCHAR	描述	否	否	授予租户账户角色权限	-	编辑
status	TINYINT	授权是否开启: 0->禁用; 1->启用	否	否	1	-	编辑
gmt_create	DATETIME	创建时间	否	否	2019-12-05T06:11:57.000+0000	-	编辑

03 产品优势

1

跨云平台计算引擎部署

解决多机房、多主体的数据安全和合规协同，助力大型集团公司或产业联盟解决复杂数据业务场景

2

多源异构的数据集成

支持离线、实时数据采集；支持多种数据类型的可视化及脚本开发；同步任务支持故障点恢复

3

数据规范定义和建模研发

结构化构建统计指标；可视化构建逻辑表模型

4

标签体系的快速构建

标签工厂模块内置标签业务方法论；提供可视化建标签能力，可快速构建企业数据标签，实现营销投放

5

数据任务统一运维异常数据监控

完善的基线配置及告警机制，保障数据任务的稳定高效；支持表级、字段级数据质量监控

6

数据资产管理和元数据盘点应用

梳理企业分散的数据信息；梳理业务信息统一数据口径；快速搭建应用洞察数据价值

1. 跨云平台计算引擎部署

数据智能应用

可信计算

数据集成

数据研发

生产运维

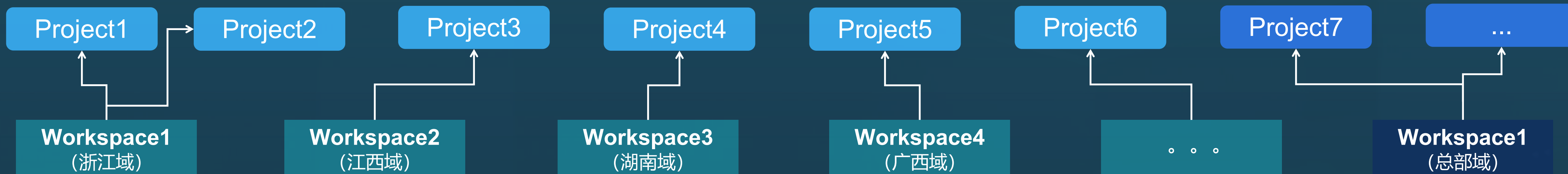
数据服务

数据治理

数据规划

营销投放

权限和审计服务



2. 多源异构的数据集成

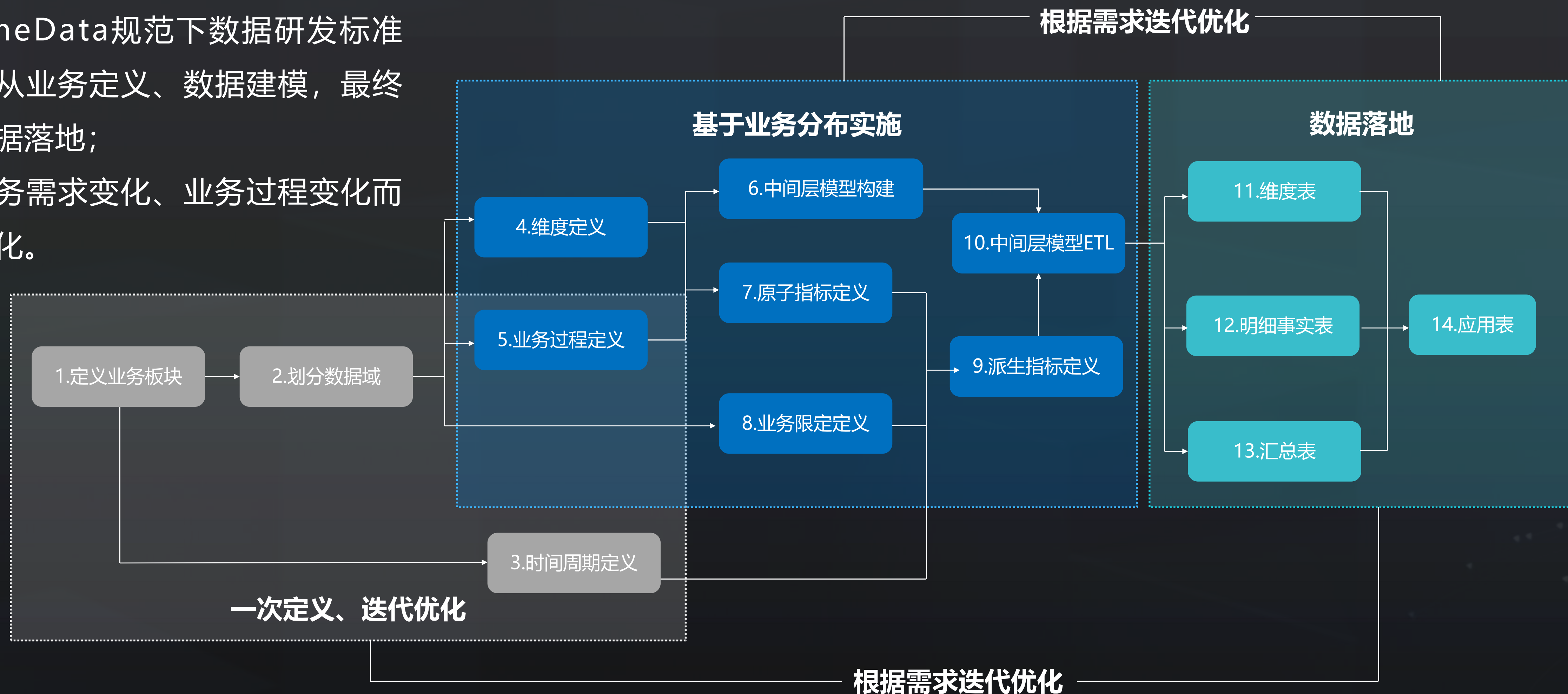
- 离线数据集成支持16种数据源类型以及文件类型数据源；支持可视化和脚本两种开发模式；可自主配置同步速率、并发数等
- 实时数据集成支持4种数据源类型；同步任务支持故障点恢复





3. 规范建模方法最佳实践

基于OneData规范下数据研发标准流程，从业务定义、数据建模，最终实现数据落地；

根据业务需求变化、业务过程变化而迭代优化。

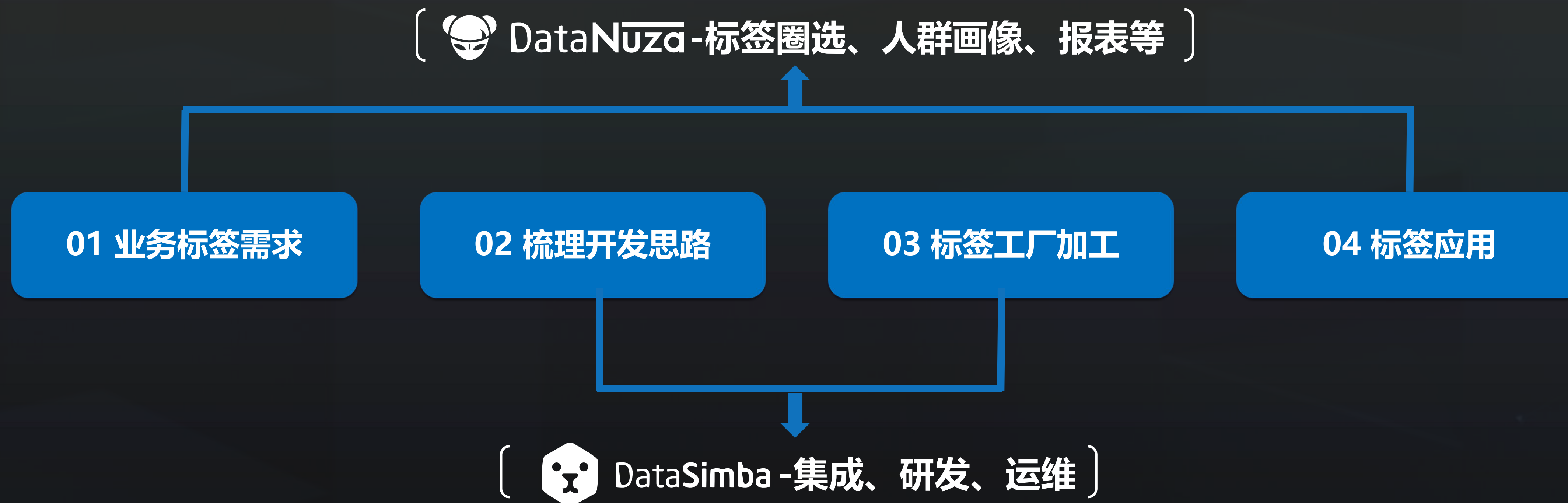


4. 标签加工应用从使用者角度划分产品边界

从使用者角色角度划分产品边界， DataSimba 标签生产 +  DataNūza 标签应用，使各角色更专注自己领域的任务，提高易用性和工作效率。

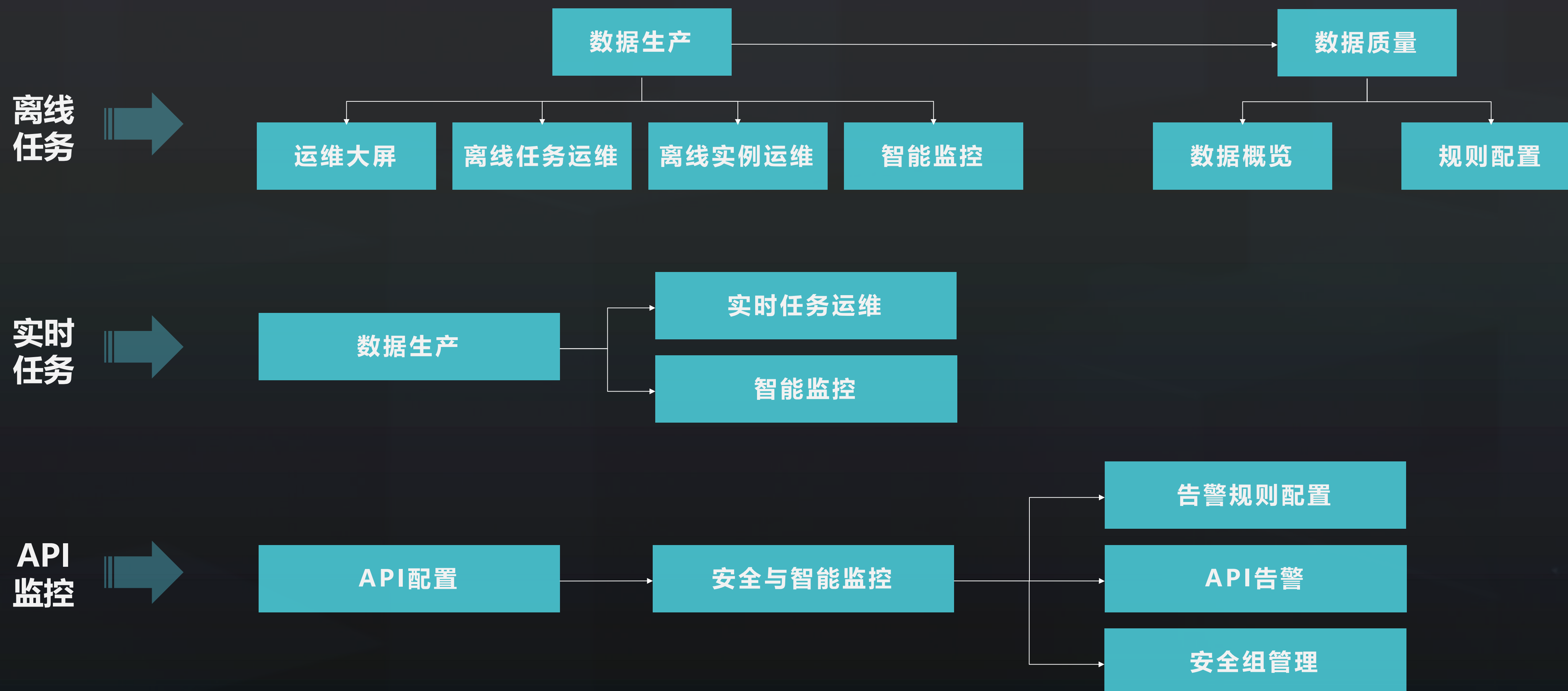
业务运营人员：确定业务目标，专注消费者分析，高效使用数据驱动精细化运营；

数据开发人员：专注标签需求涉及的数据任务，进行工程性开发，保证标签的稳定高效运行。



5. 数据任务统一运维管理

从离线任务、实时任务、API监控3个层面进行管理，确保正常调度和数据质量



6. 基于元数据的数据资产管理应用

数据治理模块包含元数据管理规划和基于元数据的数据资产管理应用。

- 元数据管理是数据资产管理的重要基础，是为获得高质量的、整合的元数据而进行的规划、实施与控制行为。元数据管理包含采集技术、业务、管理元数据，制定和管理数据标准，基于元数据的血缘分析等。
- 数据资产管理应用能够帮助企业和管理人员更进一步地洞察数据价值。基于元数据，可搭建一系列数据资产管理应用，如资产报表、资产目录、数据资产索引、数据质量核验、数据权限管控和安全审计等。



04 产品定位



DataSimba 与数据仓库、BI的差异

数据仓库的主要场景是支持管理决策和业务分析，而数据中台则是将**数据服务化**之后提供给业务系统，目标是将数据能力渗透到各个业务环节，决策分析类场景只是数据能力之一。

分类	核心能力	应用场景	计算存储	数据模型	数据治理
数据仓库 (DW) 1.0	面向业务主题的指标体系 (T+1)	报表	数据库选型: Oracle、Greenplum	ODS、DW、DM	无方法论 功能需要集成其他工具
数据仓库 (DW) 2.0	面向业务主题的指标体系 (准实时、预测类指标)	报表	Hadoop生态选型: 开源Hadoop、CDH	ODS、DWD、DWS、ADS	无方法论 功能需要集成其他工具
BI	数据可视化工具 (固定式报表及自助探索分析)	报表	部分BI工具 自带轻量级数据库	轻量级数据加工	无
数据中台	提供数据复用和服务能力的 数据资产体系	多元化场景, 包括报表、商品推荐、精准营销、个性化信息推送、销量预测等	多元化选型: Hadoop、Spark等	ODS、DWD、DWS、ADS	有方法论和产品



DataSimba 与数据湖的差异

区别	数据湖	数据中台
定义	所有企业数据统一存储	抽象数据能力等共性，形成通用数据复用与服务能力
业务场景	海量历史数据存储	多元化场景：报表、商品推荐、精准营销、个性化信息推送、销量预测等
数据类型	支持结构化数据（行与列），半结构化数据（CSV、日志、XML、JSON），非结构化数据等	支持结构化、半结构化、非结构化数据
底层技术	存储计算分类，如计算EMR和存储OSS等	多元化选型：Hadoop、Spark、Kylin、Greenplum
数据库产品	不涉及数据库选型，以HDFS等文件形式存储	主流大数据分布式：Hive、Hbase、Redis、Kudu等
数据模型	无数据模型，不进行数据处理	维度模型、雪花模型、星型模型（划分业务域、分为ODS、DIM/DWD、DWS、ADS四层）
数据治理	存储原始数据，不进行数据处理，无数据治理方法论和产品	有数据治理方法论，并基于数据治理方法论提供数据治理的产品能力
数据组织	无	面向业务场景的服务能力

业务系统提供原始业务数据，由数据中台进行数据加工、治理、计算反哺业务系统的同时，为用户运营、数据分析等模块提供数据服务。



05 场景案例

某国际奢侈品集团 | 整合全品牌数据，助推全域消费者营销

- 打通系统
- 建立标签
- 分析场景
- 统一指标

20+

275

10

150+



某国际奢侈品集团，创立于法国，旗下拥有50多个品牌，是世界最大的奢侈品集团之一。

借助数据中台，该集团打通了多个前台和后台系统，统一指标、建立消费者标签，沉淀一方数据资产。在5个首先试水的品牌中，数据中台实践获得成功，并逐步推广到旗下更多品牌中。

客户评价：

在这之前，我们对于自有的数据没有精准的感知，总是临时想要什么拉什么。对于前端数据采集和后端数据分析都是一个大难题，数据中台的建设大大缓解了这一难题，奇点云的服务也给我们留下了深刻印象。

▪ 在线化业务环比

▪ 总体业务环比

▪ 营销ROI

9倍 ↑

11% ↑

8.23+



山东得益乳业股份有限公司, 一家集生态化农业种植、规模化奶牛养殖、智能化乳品加工、现代化低温物流、数字化营销服务于一体的山东低温奶制造企业。得益乳业专注于国际流行的低温奶生产加工与研发, 赢得了千万家庭的信赖。

得益乳业以消费者数据运营为抓手, 推进数智化转型, 搭建数据中台, 沉淀一方数据资产, 并围绕消费者展开全渠道的运营和赋能。

客户评价:

在此之前, 我们缺少对消费者的了解, 且无法支持差异化营销。现在, 消费者的在线化让我们可以更贴近会员, 了解他们, 并进行精细化营销, 整体GMV也得到了巨大增长。

01 业务场景调研

针对得益现有销售场景, 如商超、自营门店、线上商城, 对其场景属性、运营目标、触点、购买模式等进行详细调研; 站在消费者角度, 模拟所有的消费场景、行为、外部影响因素及需求梯度。

02 场景上线链路设计

针对不同的业务场景, 从0到1设计不同场景的上线链路, 通过不同链路落地页和优惠券发放, 实现消费者在线化。

03 分场景差异化策略

以消费者标签为承接, 通过DataNuza的CDP实现不同人群洞察, 包括人群生命价值流转和人群画像等功能, 以此确定运营人群、优先级及运营策略

04 分场景差异化运营

根据标签人群细分进行精准化分群运营, 通过小程序、公众平台、短信等前端触点, 实现千人千面精准营销。

- 政策兑现效率
- 拉通部门数据
- 助力服务企业

2倍

20+

12万家

01 企业数据分析大屏：通过区域、街道、园区、企业等多维度数据大屏展示，协助政府自上而下数据化决策

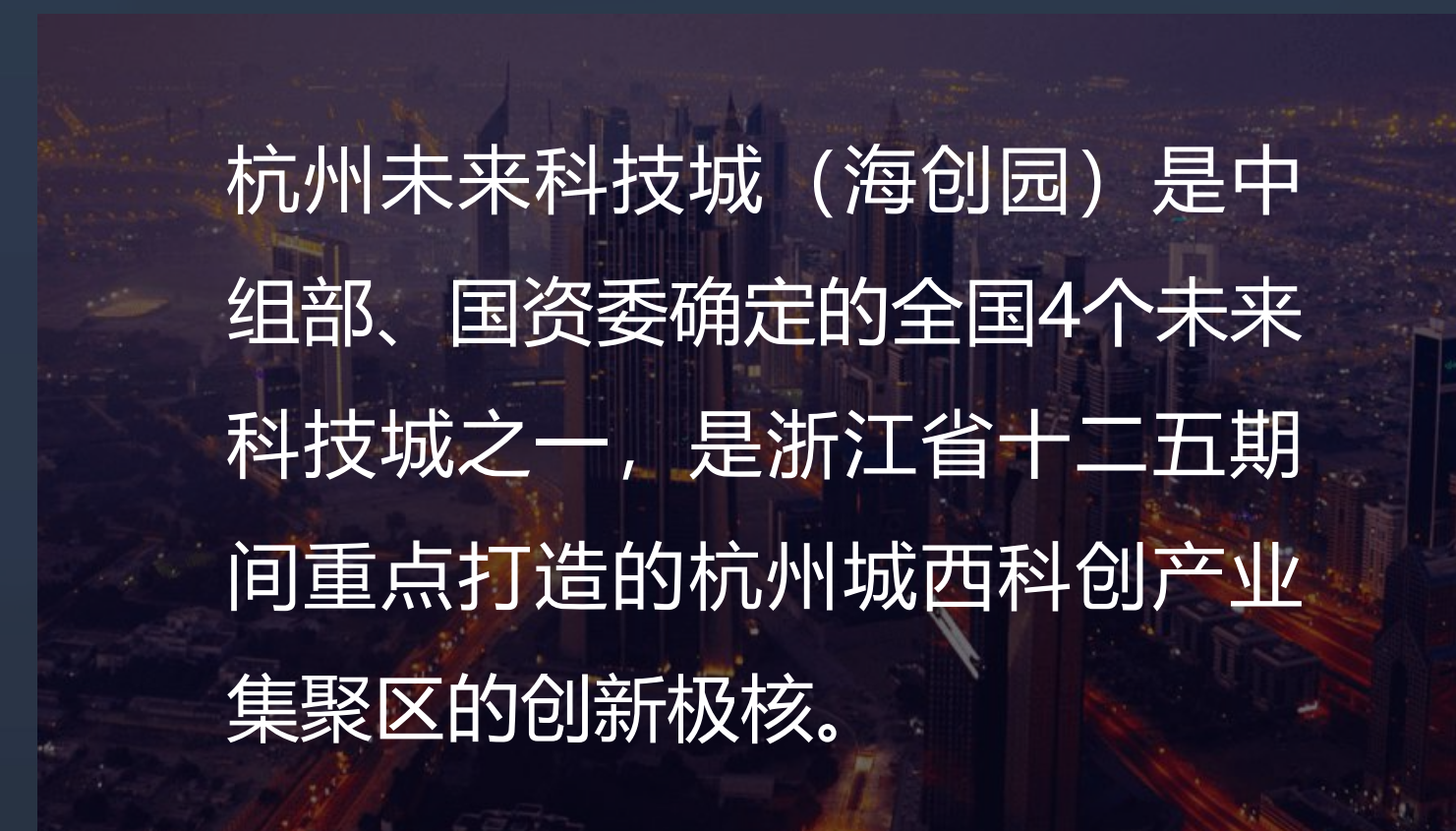
02 企业多维分析平台：从各部门出发，对部门关心的核心数据和指标层层拆解，协助部门精准管理

03 政务工作统一平台：实现对办文、办会、办事等业务工作的全流程、全要素、多视角信息化管理

04 企业服务统一门户：统一企业对外门户，实现企业项目申报、政策兑现、企业入驻等活动的统一入口

05 创业全景分析平台：清晰展现各类重点经济创新指标、人才需求状况以及综合动态榜单等数据，为政府部门提供了强有力的数据支撑

06 钉钉便携助手：整体及各园区的数据指标在移动钉钉端一目了然，方便决策者随时随地查看，且提供实时预警



杭州未来科技城（海创园）是中组部、国资委确定的全国4个未来科技城之一，是浙江省十二五期间重点打造的杭州城西科创产业集聚区的创新极核。

客户评价：

每当有申请，现在我们随报随批，不用按批处理。系统对每个环节都设置自动化了管控，例如当事项流转 to 某个部门时，必须要在若干天内处理，办理逾期、不规范审批等操作将推送警示至相关监管部门进行核实处理。

全流程在线化也便利了财务工作，转线下处理打款为线上数据留存沉淀，让台账更清晰。

 奇点云 | 让商业更智能



官网: startdt.com

电话: 0571-28121848

邮箱: business@startdt.com