

# 产品白皮书

## 数据资产管理

滴滴云

发布日期：2019-08-02

版本：V1.1

# 目录

1、适用用户.....	- 3 -
2、概述.....	- 3 -
3、产品背景.....	- 3 -
4 产品介绍.....	- 6 -
4.1 产品功能架构.....	- 6 -
4.2 产品系统架构.....	- 9 -
4.3 产品特点.....	- 10 -
4.3.1 连接.....	- 10 -
4.3.2 众包&分享&高效协同.....	- 11 -
4.3.3 机器学习.....	- 11 -
4.3.4 安全合规.....	- 12 -
4.3.5 运营.....	- 12 -
5、产品价值.....	- 12 -
5.1 企业.....	- 13 -
5.2 CDO.....	- 13 -
5.3 数据管理者.....	- 14 -
5.4 数据生产者.....	- 14 -
5.5 数据消费者.....	- 15 -
6、产品优势.....	- 15 -
6.1 按需定制.....	- 15 -
6.2 敏捷部署.....	- 16 -
7、服务保障.....	- 16 -
8、产品定价.....	- 16 -

## 1、适用用户

CDO（首席数据官）、数据管理者、数据生产者、数据消费者等从事“数据资产管理、数据应用”的用户。

## 2、概述

本文档将向您介绍“数据资产管理”产品（以下简称“产品”）的设计背景、产品的功能架构、系统架构、产品特点、产品价值、产品优势、针对性的介绍应用场景，并且能向您展现产品优势及我们提供的服务保障能力，列举出产品的定价方式供您选择。

## 3、产品背景

数据成为资产，已经是行业共识，甚至有人建议将数据计入资产负债表。但如果对比实物资产，对数据资产的管理，还处于非常原始的阶段。往往一个企业针对其数据资产类别和数量都缺乏全面了解，数据质量、数据安全、资产价值度量、资产共享等精细化管理、价值释放和持续运营则更为薄弱。

数据资产管理是现阶段推动大数据与实体经济深度融合、新旧动能转换、经济转向高质量发展阶段的重要工作内容。

- 数据资产管理已在国家、政府、行业、协会中受到高度重视，并实施多项工作，如：
  - 在政府机构改革中，设置专门的数据管理机构，已有多个省份设置了厅局级的大数据管理局，统筹推进地方“数字政府”建设；
  - 2017年7月，贵州省大数据发展领导小组办公室印发实施了《贵州省政府数据资产管理登记暂行办法》，成为全国首个出台政府数据资产管理登记办法的省份；
  - 2018年5月21日中国银行保险监督管理委员会发布《银行业金融机构数据治理指引》，明确指出银行业金融机构应当建设和优化数据治理架构，提高数据管理和数据质量控制水平，全面实现数据价值；
  - 2018年5月25日，一般数据保护条例(GDPR)正式在欧盟实施。各国对于数据跨境流动的关注则包含了数据主权、隐私保护、法律适用与管辖、乃至国际贸易规则等内容；

- 工业互联网产业联盟（AII）联合中国信息通信研究院发布了《中国工业企业数据资产管理调查报告（2018）》；

通常，数据资产管理需考虑以下几个方面以促进发挥数据价值：



- 在管理方面：管理者在规划数据文化建设时，对企业数据资产的全局构成、使用形式、使用效果都需要详细的指标输入，往往这些指标都没有被统筹起来；在组织保障上，需要多少资源、运作机制应该如何制定才能保障数据文化的落地，也需要运营指标来辅助决策，所以管理者通常需从以下几个方面的问题进行思考：
  - 数据如何被用起来？
  - 数据保值后如何增值？
  - 组织已不再满足变化所需？
  - 管理体系如何建立？
- 在治理方面：企业拥有大量的数据资产之后，由于分工不同，一般的数据生产者、数据消费者之间会随着时间推移、人员变动等因素，造成数据资产的信息成为无人维护的静态状态，数据的存储成本、检索的理解成本会越来越高。这些数据资产分布在一片数据沼泽中，难以分辨数据资产的成本、价值，更难以进行生命周期管理，甚至给数据消费者带来难以跨越的信息鸿沟；我们长期关注数据治理，并且在滴滴的长期实践中，总结出数据治理通常关注以下几个方面的问题：
  - 数据的成本如何降低？
  - 数据生命周期如何管理？
  - 数据质量低，如何保证可用？
  - 数据价值如何评估？
- 在运营方面：数据资产从被建立，到数据内容的生产、到被使用，各环节用户各自所关注的、所进行的工作重点不一致；从数据管理视角、数据生产视角、数据应用视角来看，各个视角之间的目标实现、工作重点、协作方式，不再以点对点的形式存在，而是贯穿于整个数据链路中，数据运营正是为了从以上角度来发现问题、解决问题，作用

是：数据运营会从“战略、执行、目标拆解、跟踪实现”各个阶段进行统筹，对运营目标负责。数据运营通常关注以下几个方面的问题：

- 有限的资源如何科学分配？
- 数据的关系如何互相影响？
- 如何发现最迫切的问题？
- 数据运营缺乏工具、渠道；

➤ 在使用方面：数据只有被用起来，才能发挥其应有的价值。然而当前部分的企业使用数据的情况并不乐观。根据调研统计，只有约 14%的企业数据相关的从业人员认为使用数据是方便的。数据使用是否方便，可从两个维度来判断，一是工具：是否能够具备“顺畅的、快捷的、容易完成的”数据使用场景的工具集；二是时间：是否可以快速地查找、信任、理解数据。根据调研统计，有不低于 80%的时间消耗在“查找-理解-信任”数据的过程中；这两个现状成为阻碍数据使用的最大的瓶颈。我们归纳了数据使用的几大问题点，如下所示：

- 数据孤岛亟需打破；
- 发现、理解、使用数据耗时费力；
- 知识经验无法共享、迭代；
- 沟通不畅、权责不明；
- 个人信息无法归档；
- 数据安全如何保障；

以上问题使得数据价值难以得到有效发挥。

## 4 产品介绍

“数据资产管理”产品的产品愿景是“让数据被更多、更便捷、更智能的使用起来，让数据越用越好用”。

“数据资产管理”产品以改进/解决产品背景中列举的数据资产存在的诸多问题为使命。

### 4.1 产品功能架构

首页	资产浏览 文章	数据提取 SQL提取 任务管理	数据管理 数据源接入 资产分配 资产元信息维护 PII规则维护	运营管理 角色管理 用户管理 Banner管理 公告管理	工作站	帮助中心 消息中心 账号中心
公告 Banner位 搜索词云 数据看板 运营看板 数据贡献榜 最新知识 热门资产	智能检索 数据纵览 资产详情 关注 评价 分享 标签 问答 批注 众包维护信息 知识沉淀 快速使用	智能检索 数据纵览 资产详情 关注 评价 分享 标签 问答 批注 众包维护信息 快速使用	数据源类型拓展 接入数据元信息 资产绑定负责人 负责人维护元信息 定义PII类型 设置掩码规则	数据源类型拓展 接入数据元信息 资产绑定负责人 负责人维护元信息 定义PII类型 设置掩码规则	个人贡献 有权限的 关注 足迹 批注 问答 文章 任务	帮助文档 消息管理 密码管理

➤ 产品所提供能力如下：

- 数据资产接入能力；
- 快速查找-理解数据的能力；
- 使用数据的工具能力；
- 数据安全合规的基础保障能力；
- 为数据管理者提供的贯彻数据文化、数据政策的贯宣能力；
- 为数据消费者提供的工作站，一站式个人数据、信息管理能力；
- 为数据生产者提供的的数据质量维护能力；
- 为数据运营者提供的工具能力；

#### 数据资产接入能力

产品预置了常规的数据类型的 Adapter，当前支持了“ORACLE、MYSQL、HIVE、TABLEAU 报表、特征”的类型，这些类型的数据可通过配置的方式自助接入数据；同时，我们也提供快速适配接入其他数据类型的能力，如文档、其他类型数据库等。

## 快速查找-理解数据的能力

使用数据时，一般会经过“发现—理解—信任—使用”这 4 个阶段，而在实际的工作中，这 4 个阶段往往要耗费大量的时间精力，尤其是在查找数据与理解数据。因此，产品围绕“发现—理解—信任—使用”4 个阶段，着重打造相应的产品能力。

### [查找]

产品首先要做的就是打破“数据孤岛”，接入数据后，通过不断训练的智能搜索引擎，一站式快速检索全部资产，提高查找效率。

### [理解]

为了让数据更好地被理解，产品从以下方面进行了能力的覆盖：

- 强调资产到人，负责人对资产的信息质量负责，包含技术信息、业务信息；
- 采用“众包”的产品理念，资产的描述信息通过消费者的不断完善来提升质量；
- 引入用户评价体系，数据好不好用，质量高不高，消费者最有发言权；
- 引入“红黄绿”标注功能，对资产的警告、提醒、推荐信息，全局维护全局受益；
- 私有的标签库，更符合个人的标签体系，易于管理；
- 细化到资产的批注功能，“资产-信息-知识”无缝连接，提升知识沉淀能力；

### [信任]

在长期实践中，我们发现数据消费者在使用数据时，会首先对数据进行探查，通过探查的结论，来判断数据内容、数据质量是否符合预期，是否能够满足需求，因此，产品提供了对数据进行探查的相应能力，如：枚举值、数值分布（最大值、最小值、平均值）、数值 count 数、数值 distinct count 数，空值数等统计，并辅以图表进行呈现。

### [使用]

在数据使用效率方面：我们充分利用机器学习的能力，持续分析用户数据使用行为，通过对使用场景的识别，对使用数据的客观规律的提炼，让数据更懂用户：

- 对数据之间关系的智能分析，提供智能数据推荐服务；
- 对数据使用规律的智能学习，让 SQL IDE 更加智能，提高 SQL 代码效率；

## 使用数据的工具能力

在数据使用工具方面：产品提供了基于 SQL IDE 的提取工具，数据消费者可按需通过 SQL 代码实现对数据提取。而对于分析的场景，产品集成了 tableau 的可视化分析能力，可快速创建 tableau 数据源，创建 tableau 可视化报表，为企业快速部署可视化工具提供了快捷的选项。

除了产品提供的通用能力，我们也支持接入企业的数据提取工具产品、可视化分析产品的能力，可按需完成服务能力的接入，从而打造数据查找-使用的交互闭环。

## 数据安全合规的基础保障能力

数据安全是所有企业共同面临的挑战，产品在数据安全合规方面，也提供了基础的保障能力：

- 数据安全分级：支持企业根据实际管理需要定义数据安全等级；
- PII 定义：预置 GB/T35273-2017 的 PII 信息管理和规范，也支持企业进行扩展；
- PII 数据掩码：对于 PII 数据，可以在产品中设置对于 value 的掩码规则，防止泄露；
- 安全审计：完整地记录不同安全等级的数据资产的使用记录，便于安全审计，并可以根据企业要求定制安全预警服务；

### **为数据管理者提供的贯彻数据文化、数据政策的贯宣能力**

产品作为数据资产的全局维护、管理、发现的中心，适合作为贯彻数据文化、数据政策的贯宣工具，我们提供信息内容模块丰富的首页，提供了诸如公告栏、活动广告 banner、运营指标、数据贡献榜、搜索洞见等模块，让企业的文化得以展现、推广。

### **为数据消费者提供的工作站，一站式数据、信息管理能力**

本着以用户为中心的原则，我们提供了“工作站”，为数据消费者提供一个管理个人数据、信息的空间，包含了用户的“有权限的、关注、足迹、文章、问答、批注、任务”信息。

### **为数据管理者、生产者提供的数据质量维护能力**

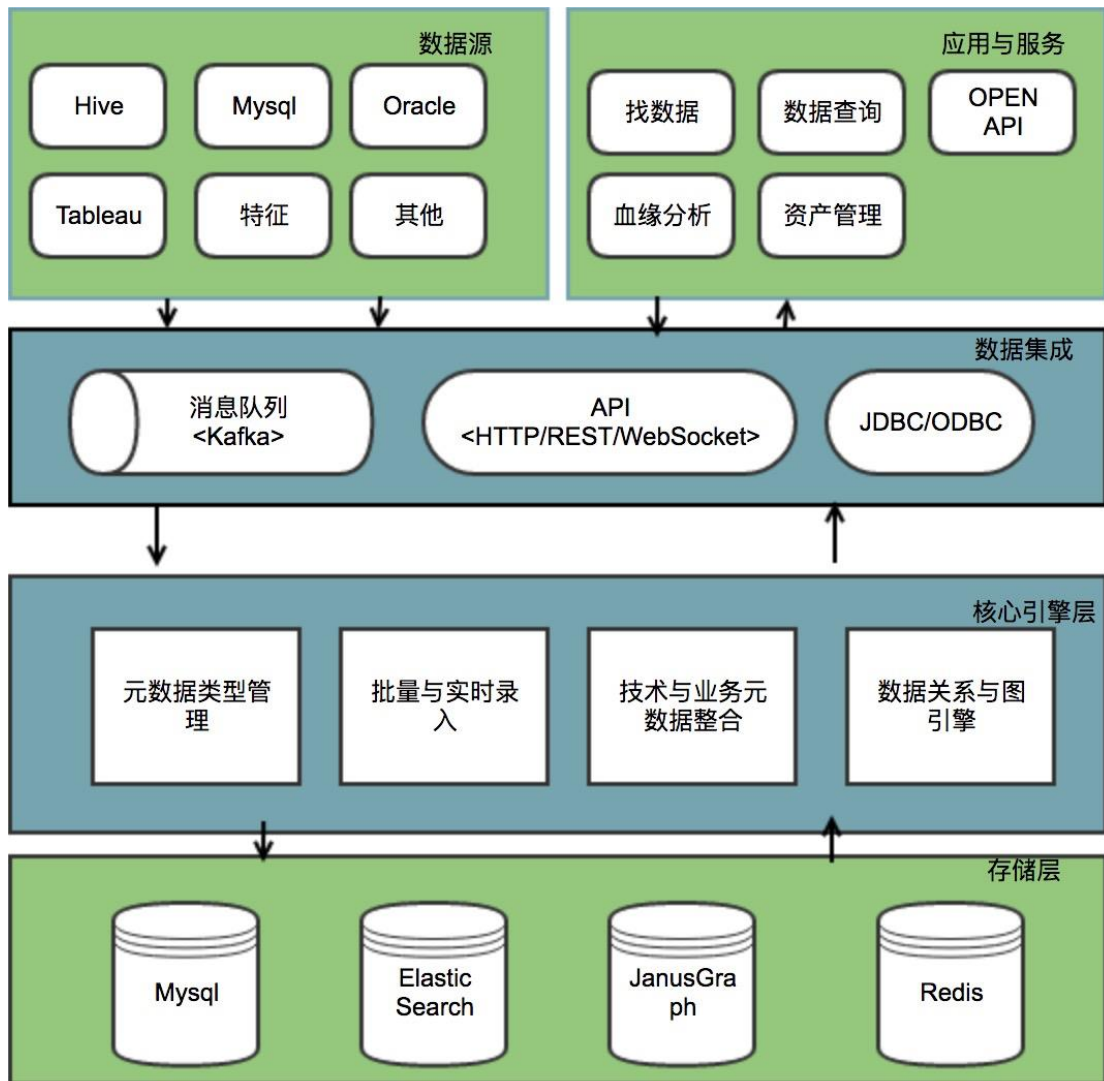
从资产被接入之后的“负责人维护、安全等级维护、数据元信息维护”，都可以通过“数据管理”模块来完成，这一过程可进行集中管理维护，使得数据元信息完成从 0 到 1 的初始化。

### **为数据运营者提供的工具能力**

数据文化的落地，落地效果，可通过运营工具来支撑，如，对数据质量的运营、数据信息的维护、对数据贡献的奖惩、对促进数据相关知识的贡献源，都可通过文章、贡献榜、组织活动进行推广，产品提供基础的运营能力，同时我们也提供按需定制运营能力的支持。



## 4.2 产品系统架构



- **数据源**

支持多种数据源的元数据和数据查询，包括 Hive、Mysql、Oracle、Tableau、特征和其他类型。

- **应用与服务**

为用户提供多种功能的应用与服务，包括找数据、血缘分析、资产管理和数据查询等，还支持 OPEN API，供第三方集成使用。

- **数据集成**

数据集成层支持多种集成方式，包括消息队列、HTTP Restful / Websocket API、和 JDBC/ODBC 方式，通过多种方式将数据源的数据整合进入核心引擎，同时对上层应用和服务提供支持。

## ● 核心引擎

核心引擎层包括多种组件，支持所有的产品能力，包括多种类型元数据的 schema 管理、元数据的批量与实时录入、采集的技术元数据与用户产生的业务元数据的整合、数据之间血缘和引用等关系的建立与图数据的存储与查询引擎等，通过这些核心引擎来支持产品能力。

## ● 存储层

存储层保存着产品所有的数据，并针对不同的使用场景分别采用不同的存储引擎，包括 Mysql、ElasticSearch、图数据裤 JanusGraph、和 Redis。

## 4.3 产品特点

### 4.3.1 连接

实践过程中，通过对资产相关内容的抽象，我们分离出围绕资产的“人、信息、知识、智慧”，这些内容虽然在物理空间中彼此独立，但却始终以资产作为核心，经过沉淀后将成成为资产最富有活力的衍生内容。



- 通过资产，连接“人-信息-知识-智慧”，多维度的描绘资产的全景，让资产易于被理解、被信任，减少使用数据的中间过程；
- 连接后的信息、知识，通过机器学习，挖掘用户行为、识别使用的场景差别，从而更懂用户，更能辅助用户；

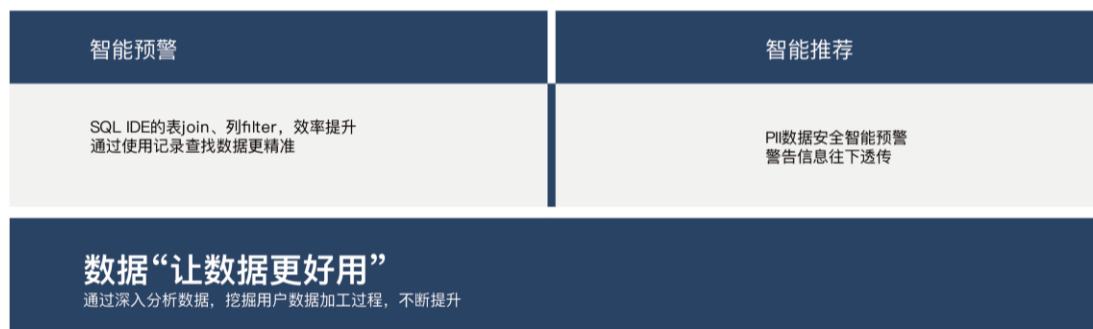
### 4.3.2 众包&分享&高效协同

数据消费者，同时也是数据生产者，通过众包，号召持续完善数据元信息、数据记录、解答提问、评价资产、使知识更能真实地呈现；

数据需要进行流通起来，好的数据资产应该被更多的分享使用，并对数据生产者的付出进行奖励；

经过实践，使用数据过程存在大量的需要及时沟通协作的场景，故通过 IM、电话、邮件等方式进行高效协同是十分必要的；

### 4.3.3 机器学习



➤ 得益于围绕资产进行的“人-信息-知识-智慧”的连接，并通过持续的算法改进，我们进行了以下成功实践：

- **效率提升**：挖掘数据使用的客观规律，识别数据使用的场景规律，并提供高质量的智能推荐，提升数据使用效率；
- **辅助管理**：将一些关键的、需进行大量人工干预的环节，通过不断完善的机器训练成果，对数据安全的智能预警、报警信息的自动透传方面进行了能力提升，提升工作效率；
- **规范数据标准**：生产的数据由于格式、标准不统一等现象，对数据使用造成阻碍，实际使用时需频繁进行人工处理以达到使用要求。而这些高频的重复动作，可通过数据整理引擎，从数据结构层进行全局统一，建立数据使用标准；

#### 4.3.4 安全合规

数据安全策略	数据安全措施	数据安全审计
资产数据安全分级 预置国标PII类型及名称	PII数据掩码 自定义PII类型及掩码翻译 数据权限管理	访问日志 使用日志 敏感信息使用预警

- 确保数据获取和使用合法合规
- 为数据价值的充分挖掘提供了安全可靠的环境。

数据安全合规是企业的责任，除了做好数据在存储、交换、窃取的安全合规之外，还需充分考虑数据共享时的脱敏。产品预置了 GB/T35273-2017 的 PII 信息管理和规范，并提供自定义数据掩码的能力，帮助企业在数据共享这一应用场景中实现数据的安全合规。

#### 4.3.5 运营

数据资产管理发展到现在，已不局限于“管理”的范畴。对资产进行管理，本质上是为了让数据资产发挥价值。这个目标需要从顶层设计到运营辅助进行支撑，数据运营是产品持续强调的产品能力。

数据价值为目标	用户为中心	业务价值为导向
推优点、补短板、 奖贡献、罚不力	分角色、分场景、 分粘性	找痛点、挖需求

**价值评估 共享 流通**

- 对数据价值进行评估，需求合理的方式管理内部数据和提供数据服务
- 促进共享有助于消除数据孤岛，打通各系统的数据，促进数据价值释放
- 推进数据流通有利于让数据服务能力得到提升，从而实现增值
- 运营需要组织保障，人员保障

- 对数据价值进行评估，寻求合理的方式管理内部数据和提供数据服务；
- 促进共享有助于消除数据孤岛，打通各系统的数据，促进数据价值释放；
- 推进数据流通有利于让数据服务能力得到提升，从而实现数据保值、增值；
- 运营需要组织保障，人员保障；

## 5、产品价值

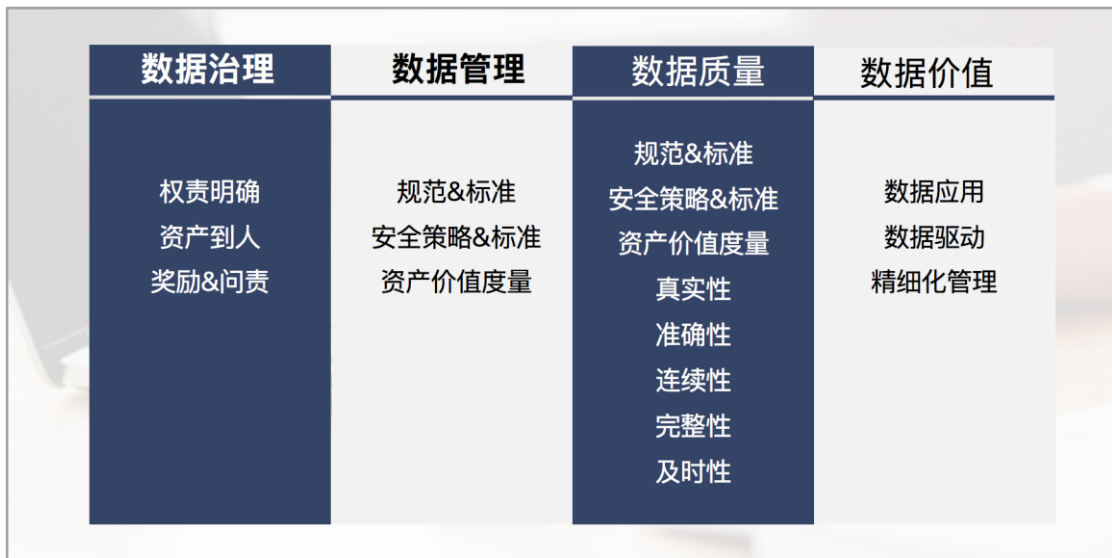
我们致力于为 CDO、数据管理者、数据生产者、数据消费者等从事“数据资产管理、数

据应用”的用户提供更聚焦的服务，实现“让数据被更多、更便捷、更智能的使用起来，让数据越用越好用”的产品愿景。

## 5.1 企业

目前企业的数据安全合规已经进入严监管、强监管时代。因此，各企业必须快速加强数据资产管理、数据治理体系建设，在数据的完整性、准确性、一致性、时效性以及安全管理、制度建设等方面查漏补缺，满足监管要求。

我们将产品能力按照以下 4 个方面进行赋能，助力企业：



通过产品，建立企业的标准体系和配套管理流程和工具，形成能够准确表达业务的、企业统一的数据语言，成为连接基础性数据工作和数据应用的桥梁，有效解决“数据不够用”、“数据不好用”的典型问题。

## 5.2 CDO

从顶层视角，为 CDO 提供以下信息：



- 接入企业数据资产，总览企业数据资产的构成；
- 监控数据质量，制定数据治理、数据管理的政策；
- 贯宣落实企业的的文化；

### 5.3 数据管理者

为数据管理者关注的核心指标提供识别、拆解、落实的抓手：

有什么资产	使用情况如何	数据质量如何衡量、提升	资源、成本如何优化	数据安全如何保障
提供灵活的数据源接入，绘制统一数据视图	数据只要被使用，都会被统计分析，轻松识别核心、关键资产	基于数据血缘、用户使用记录、数据生命周期的资源、成本优化	资产到人，数据内容完善质量自动分析，配合运营工具，提升质量	预置GB/T 35273-2017标准的PII信息标记、数据掩码，支持自定义

- 提供灵活的数据源接入，绘制统一数据视图；
- 每条数据使用记录都会被统计分析，轻松识别核心、关键资产；
- 基于数据血缘、用户使用记录、数据生命周期的资源进行成本优化；
- 资产到人，数据内容完善质量自动分析，配合运营工具，提升质量；
- 数据安全合规的应用、信息审计；

### 5.4 数据生产者

数据生产者作为提供数据内容、对资产质量负责的重要群体，我们总结得出以下 3 大问题并提出解决方案：

和数据消费者高效协同	贡献如何被感知	经验需要被沉淀
时效要求高：IM，电话、邮件，多种方式选择 时效要求低：发表问答、评论	通过贡献体系，维护数据元信息、回复、被点赞、发表文章，都会体现到个人贡献 资产的用户5星评价体系	常见咨询，发表文章，与更多的人分享 回复问答，让其他人学习，提升

#### 贡献如何被感知

- 通过贡献体系，维护数据元信息、回复、被点赞、发表文章，都会在个人贡献中得到体现；
- 资产的用户 5 星评价体系；

#### 经验需要被沉淀

- 常见咨询，发表文章，与更多人分享；
- 回复问答，与他人共同学习、共同提升；

#### 和数据消费者高效协同

- 时效要求高：IM，电话、邮件，多种方式选择；

- 时效要求低：发表问答、评论；

## 5.5 数据消费者

耗时占比-80%	认为使用数据不便的占比-86%
用户用于“查找-理解-信任”数据花费的时间与完成数据使用时间的占比	只有14%的用户认为使用数据是方便的

- 统一浏览，打破数据孤岛，快速精确查找所需数据；
- 支持数据的概要信息浏览，理解数据快人一步；
- 方便地建立数据消费者与生产者的沟通协作，高效答疑；
- 完善数据使用记录，学习数据使用经验；
- 强大的个人批注，信息与资产无缝连接，专属于个人的使用笔记；
- 便捷的数据实用工具，SQL 查询、tableau 可视化分析一键直达；

## 6、产品优势

### 6.1 按需定制

根据企业环境，提供按需定制的能力，无需打包为其他附庸付费

产品模块	说明
数据目录	不依赖其他模块的基础服务，最小可用产品，含（首页、数据管理、数据浏览、工作站、运营管理、用户指南、消息中心、账号中心）
定制数据源	根据企业数据类型，快速完成接入，这一过程最快仅需天级实现
可视化 (独立模块)	支持 tableau 可视化工具，并可对接企业专有的可视化工具，完善数据使用渠道
SQL 查询 (独立模块)	提供 SQL IDE 工具，包含 SQL 代码查询数据，任务管理
文章 (独立模块)	数据内容社区模块，沉淀数据知识，数据运营的一个载体

Notebook (独立模块)	创建/分享“SQL 数据查询、结果可视化、文档”的高阶工具
安全审计 (独立模块)	管理敏感数据使用记录、使用详情

## 6.2 敏捷部署

产品支持私有化部署和云部署，企业可按需选择：

云部署	私有化部署
按需订购，备份、维护更简便 更低的投入成本，灵活的资源扩展能力 无需专业IT运维，更少的运维成本	投入人员、资金、硬件，前期投入较大 运维成本高 需长远规划，投入资源以扩展规模

## 7、服务保障

依托滴滴云完善的售后服务体系，我们以“One of top Service in the World”作为服务目标，提供“一年免费标准版维保服务、付费高级版维保服务”供企业按需选择：

维保	增值
<p><b>一年免费标准版维保服务</b></p> <p><b>服务期限</b> 一年 5x8</p> <p><b>服务项目</b> Bug修复 技术咨询服务</p> <p><b>服务收费标准</b> 免费</p> <p><b>服务响应级别</b> 初始服务响应时间 60min 初次故障分析回复时间 90min 初次解决方案回复时间 2个工作日</p>	<p><b>付费高级版维保服务</b></p> <p><b>服务期限</b> 按需购买</p> <p><b>服务项目</b> 数据应用流程咨询服务 数据安全治理咨询服务 运营管理咨询服务 数据治理服务咨询</p> <p><b>服务收费标准</b> 1万元/人*天，10人*天为最小购买单位</p> <p><b>服务响应级别</b> 下一工作日到场</p>

## 8、产品定价

产品定价，请联系：[dataprod.didicloud@didiglobal.com](mailto:dataprod.didicloud@didiglobal.com)

为确保高效响应您的需求，邮件中，请确保提供以下信息：

公司全称：



需求部分：

联系电话：

联系人姓名：

滴滴云