

大数据智能体工作台 DataBuddy

产品简介



腾讯云

【 版权声明 】

©2013–2026 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【 商标声明 】



及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【 服务声明 】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【 联系我们 】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或 95716。

文档目录

产品简介

产品概述

产品优势

产品架构

产品功能

应用场景

产品简介

产品概述

最近更新时间：2026-08-12 16:15:00

产品定义

大数据智能体工作台 DataBuddy 是腾讯云推出的 **Agent 原生（Agent-Native）**、**全托管**的 Data + AI 一体化数据智能平台。DataBuddy 融合数据计算与 AI 智能体能力，通过统一元数据与语义层，为企业提供从数据接入、数据工程、数据科学、数据分析到数据治理的**端到端全链路能力**。

DataBuddy 让大数据平台从“人操作工具”进化为“AI 工作、人把关”——用户通过自然语言即可完成数据接入、ETL 构建、指标开发、异动归因、可视化看板等复杂工作，降低数据平台的使用门槛与运维成本。

主要功能

DataBuddy 覆盖数据全生命周期，主要能力可归纳为“**三大 Agent、DIOps、计算存储底座、一套治理体系**”：

三大 Agent（面向业务场景）

Agent	面向角色	核心能力
工程 Agent (DataBuddy-Engineering)	数据工程师	数据接入、数仓分层方案设计、ETL 任务生成与编排、工作流调度、智能运维
治理 Agent (DataBuddy-Governance)	数据管理员 / 治理团队	敏感字段 AI 识别与分级分类、质量规则智能推荐、数仓质量诊断、元数据补齐、根因诊断
分析 Agent (DataBuddy-Analysis)	分析师 / 业务人员	自然语言问数、分析报告生成、指标波动归因、AI 搭建可视化看板、主动推送

DIOps · 一体化工程能力

DIOps（Data Intelligence Ops）是 DataBuddy 面向“数据 + AI”全生命周期的一体化工程体系，涵盖 **DataOps、MLOps、AIOps** 三大 Ops 能力：

Ops 能力	主要模块	核心价值
Data	Studio（开发）、Workflow（编排调度）、	大数据全链路开发运维闭环，数据、代

Ops	CI/CD（发布）、OPS（运维监控）	码、配置版本化
MLOps	特征管理、模型实验、模型管理、模型服务、AI Gateway	机器学习模型全生命周期管理，端到端模型上线周期从 30 天缩短至 7 天
AIOps	AI 智能诊断、根因分析、自动修复、智能运维	从“人工排障”到“AI 主动诊断”，60%+ 问题由 AI 自愈

计算与存储底座

- **数据湖计算**：承载 SQL、Python、Notebook 计算负载，支持 CU（Compute Unit，计算单元，1 CU ≈ 1 vCPU + 4 GB Memory）弹性资源
- **湖仓存储**：托管存储层，承载表、Volume 等全模态数据

一套治理体系

- **统一语义层 Unity Semantics**：本体建模、指标、维度、语义模型，降低 Agent 幻觉风险
- **数据质量 Data Quality**：全链路质量规则、异常检测、数据剖析
- **多模态数据目录 TCCatalog**：统管结构化数据、非结构化数据、ML 模型、外部数据源
- **Agent 安全防护**：提示注入检测、OBO（On Behalf Of，以用户身份代理执行）权限强约束、高危 SQL 拦截、“三要素”精准拦截

产品架构

DataBuddy 采用**交互层、平台层、引擎层**三层架构（三层一体化融合、全托管产品）：

- **交互层**：Agent（工程、分析、治理）、GUI（全功能可视化）、Notebook（交互式分析、数据科学）
- **平台层**：统一语义、Data 平台（集成、开发、调度）、AI 平台（ML、GenAI、推理）、一体化治理（数据治理 + AI 治理）
- **引擎层**：数据湖计算、湖仓存储、Data + AI 一体化、全模态支持

产品优势

最近更新时间：2026-08-12 16:15:00

Agent-Native 原生架构

DataBuddy 不是“传统平台外挂 AI 插件”的拼装式产品，而是从底层引擎到上层交互都为 Agent 场景重新设计：

- **三大内置 Agent 开箱即用**：工程 Agent、治理 Agent、分析 Agent 深度绑定腾讯云大数据生态，面向大数据领域深度优化的 Skill 体系、专属知识库、上下文优化与持续记忆
- **Agent 双向防护**：提示注入检测、OBO 权限强绑定、高危 SQL 拦截、“三要素”精准拦截，覆盖 OWASP LLM Top10 全部风险
- **持续进化飞轮**：每次使用都在沉淀知识和记忆，越用越好，模型红利持续加持到数据平台

Data + AI 一体化平台

一份数据、一套计算，同时服务数据分析与 AI 应用：

- **统一 DAG（Directed Acyclic Graph，有向无环图）混合编排**：大数据任务与 AI 任务（模型训练、推理、特征工程）在同一工作流里编排，血缘贯通
- **MLOps（Machine Learning Operations，机器学习工程运维）全生命周期**：MLflow 跟踪、AutoML、大模型实验、特征管理，覆盖机器学习全生命周期：训练、评估、部署、监控
- **成本优化**：CPU、GPU 算力复用，避免数据在两套系统间来回流动
- **上线周期**：端到端模型上线周期从 30 天缩短到 7 天，特征复用率提升 1.5 倍，开发效率提升到 2 倍

统一语义层 Unity Semantics

LLM 时代的瓶颈不在模型，而在于模型对业务的理解会更好一些。DataBuddy 通过统一语义层解决 NL2SQL（Natural Language to SQL，自然语言转 SQL）的三大幻觉陷阱：

幻觉陷阱	无语义层表现	语义层解决方案
指标歧义	“收入”有 8 种口径，Agent 随机选	指标、实体、业务口径标准化
JOIN 错乱	关联键靠猜，SQL 跑通但数据错	语义模型固化关联关系
查询条件乱写	“上季度”、“华东区”靠猜字段	本体建模明确维度绑定

效果对比：

- **无语义层**：NL2SQL 准确率 60% ~ 70%
- **完整语义层**：NL2SQL 准确率 90%+

配合治理 Agent 持续更新校验（语义不过期），“缺语义会导致幻觉、缺治理会导致过期”，同时解决这两类问题，正是 DataBuddy 的护城河。

多模态治理 · 治理对象全域化

TCCatalog (Tencent Cloud Catalog, 腾讯云统一数据目录) 一个入口统管四类资产:

- **Table Catalog:** 表、视图、函数
- **Volume Catalog:** 日志、图像、PDF、视频、音频
- **Model Catalog:** ML 模型、版本管理
- **External Catalog:** 外部数据源镜像, 免 ETL (Extract-Transform-Load, 数据抽取转换加载) 即查

配套能力:

- 端到端数据血缘: 自动采集 SQL 与集成任务, 模型血缘覆盖训练到推理全链路
- 数据质量: 新鲜度、完整度、一致性全链路监控, 异常自动检测
- 敏感数据识别: AI 自动打标分级, 一键审核确认
- 细粒度权限: Catalog、Schema、Object 三层级联

总拥有成本 (TCO) 显著降低

DataBuddy 一体化架构降低架构成本, 具体体现:

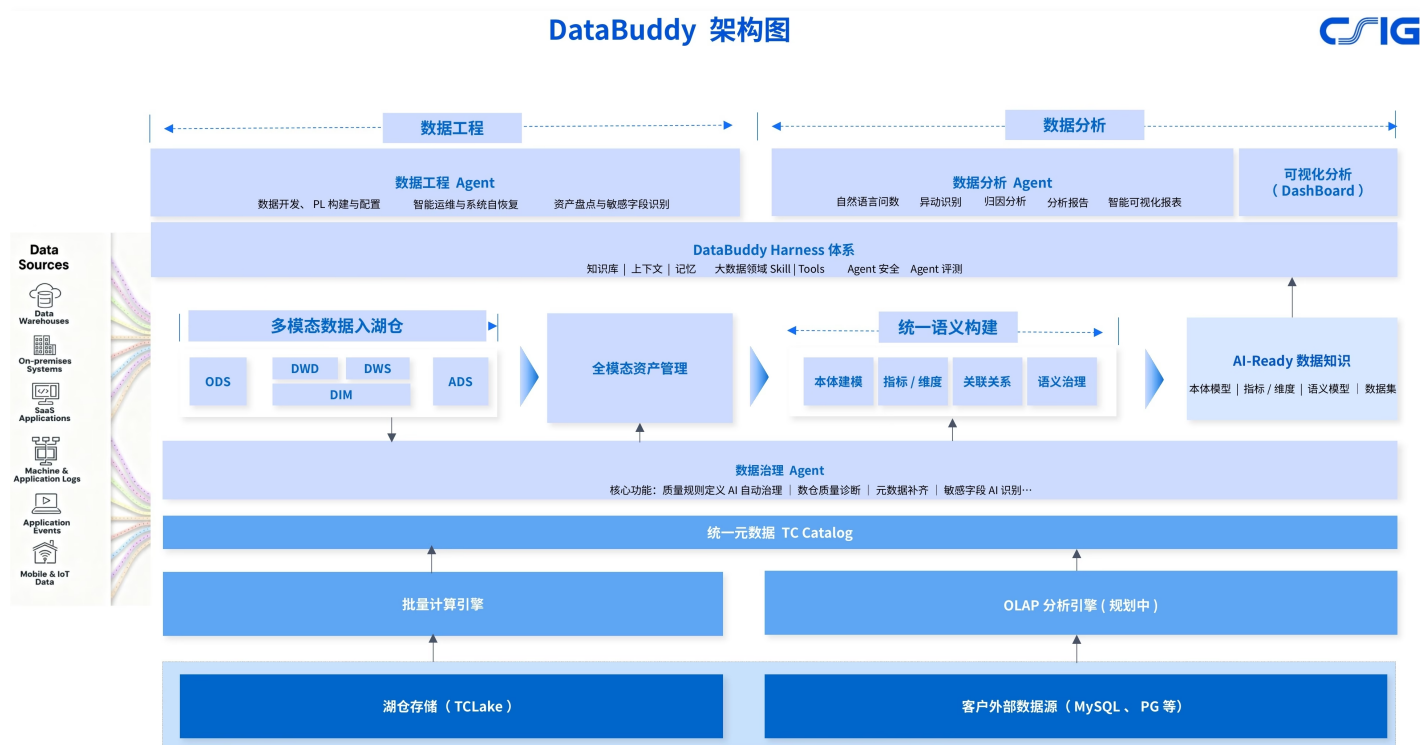
- **免版本费:** 专业版、企业版的差异仅体现在计费项单价递增——客户“零门槛”开通, 先用起来再随规模付费
- **按量、包月双计费:** 以 CU 为统一计量单位, 波峰用按量、稳态用包月, 客户可自主优化成本
- **AI 能力独立商业化:** Agent 订阅 198 元 / 人 / 月起, 积分包 5 档按需购买, 用量弹性扩容
- **云原生一体化, 隐藏复杂度:** 基础资源打包进 CU、GU (GPU Unit, GPU 计算单元), 统一订单出口, 客户无需拼装多方账单

产品架构

最近更新时间：2026-08-12 16:15:00

DataBuddy 采用 **Agent 原生、Data + AI 一体化、全模态支持** 的一体化产品架构。整个产品由上至下融合 **业务交付层（Agent + 可视化分析）、能力平台层（数据资产、语义、治理）、引擎与存储层** 三个层次，围绕企业数据从“接入 → 建仓 → 语义化 → 消费 → 治理”的全链路闭环设计。

架构总览



说明

上图对应 DataBuddy 端到端的核心链路：**多模态数据入湖仓 → 全模态资产管理 → 统一语义构建 → AI-Ready 数据知识 → Agent 或可视化分析消费**。全流程由 **DataBuddy Harness 体系** 提供 Agent 底座能力，由 **数据治理 Agent** 横向守护，由 **统一元数据 TC Catalog（Tencent Cloud Catalog，腾讯云统一元数据目录）** 提供统一目录视图，最终落地在托管的湖仓与计算引擎之上。

一、数据源接入范围

DataBuddy 支持企业内外部各类数据源的接入，覆盖以下类型：

- **数据仓库（Data Warehouses）**
- **本地系统（On-premises Systems）**
- **SaaS 应用（SaaS Applications）**
- **机器与应用日志（Machine & Application Logs）**

- 应用事件（Application Events）
- 移动与 IoT 数据（Mobile & IoT Data）

说明：
涵盖关系型数据库、数仓、湖仓、NoSQL、KV、消息队列、流、文件、对象存储、API、多模态等国内外数据源，提供 5 大接入方案：**离线同步、实时同步、文件到 Table、文件到 Volume、批量到 Volume。**

二、业务交付层：Agent 原生、面向“对话即交付”

DataBuddy 遵循 “Agent 为主、可视化为辅” 的产品逻辑：Agent 生成的代码、任务、语义模型、看板均沉淀到统一元数据 TC Catalog，可通过 GUI 或 Notebook 进行代码化、图形化二次编排。

三大内置 Agent

场景	交付 Agent	关键能力
数据工程	数据工程 Agent	数据开发、PL（Pipeline，数据流水线） workflow 构建与配置、智能运维与系统自恢复
数据分析	数据分析 Agent	自然语言问数、异动识别、归因分析、分析报告、智能可视化报表
数据治理	数据治理 Agent	质量规则定义 AI 自动治理、数仓质量诊断、元数据补齐、资产盘点与敏感字段 AI 识别

可视化分析（DashBoard）

DashBoard 是数据分析域的独立消费入口，直接从 AI-Ready 数据知识 取数：

- 承载 Agent 产出的看板、报表、探索结果。
- 支持业务人员在图形化界面上自助创建、调整仪表盘。
- 与分析 Agent 双通道协同：Agent 自动生成看板后，可在 DashBoard 中二次编辑。

三、DataBuddy Harness 体系：Agent 原生底座

Harness 体系是三大 Agent 稳定“对话即交付”的核心底座，跨越工程、分析、治理三条业务线，统一提供以下能力：

- **知识库：**企业级知识注入，包括业务本体、数据资产、SOP（Standard Operating Procedure，标准作业流程）等。
- **上下文：**多轮对话上下文管理、跨会话状态保持。
- **记忆：**Agent 长期记忆持久化，沉淀用户偏好与历史决策。
- **大数据领域 Skill、Tools：**内置面向数据工程、语义建模、治理、分析等场景的领域 Skill 与工具集。

- **Agent 安全**：双向 Guardrails（护栏机制，输入侧提示注入检测 + 输出侧结果防护）、OBO（On Behalf Of，以用户身份代理执行）权限强绑定、高危 SQL 拦截、意图校验，覆盖 OWASP LLM Top 10 全部风险。
- **Agent 评测**：Skill 与 Tool 调用轨迹追踪、任务成功率、Token 消耗、语义还原度等多维评测。

四、能力平台层

4.1 多模态数据入湖仓

DataBuddy 在湖仓内部完成多层建仓，形成标准数仓分层结构：

- **ODS（Operational Data Store，操作数据层）**：承载入湖后的原始明细。
- **DWD（Data Warehouse Detail，明细数据层）**：完成清洗、标准化。
- **DWS（Data Warehouse Summary，汇总数据层）**：构建主题汇总宽表。
- **DIM（Dimension，维度层）**：抽取公共维度供上层复用。
- **ADS（Application Data Store，应用数据层）**：直接服务于分析、报表、Agent 消费。

多模态支持：结构化数据落入 Table，非结构化、半结构化数据（图片、视频、文档、音频、日志）落入 Volume，一套建仓分层同时覆盖。

AI 加持：数据工程 Agent 支持自然语言创建接入与建仓任务、AI 智能问答、任务异常 AI 根因定位。

4.2 全模态资产管理

DataBuddy 通过 **TC Catalog** 提供统一的多模态资产视图，让“结构化、非结构化、模型、外部镜像”共享一套元数据、权限与血缘：

资产类型	说明
Table Catalog	表、视图、函数（结构化数据）
Volume Catalog	日志、图像、PDF、音视频等非结构化数据
Model Catalog	机器学习模型、版本管理、血缘
External Catalog	外部数据源镜像（免 ETL 即查）

4.3 统一语义构建（Unity Semantics）

统一语义层是 DataBuddy “AI-Ready” 的关键——通过标准化的本体建模与指标、维度体系，消除口径歧义：

- **本体建模（Ontology，业务本体建模）**：标准化实体、属性、业务逻辑，构建业务本体模型。
- **指标、维度**：企业级“语义知识图谱”，统一指标口径与维度定义。
- **关联关系**：跨主题域实体关联、属性映射。
- **语义治理**：与数据治理 Agent 联动，实现 口径校验、语义可迭代。

4.4 AI-Ready 数据知识

语义构建的产物——**本体模型、指标、维度、语义模型、数据集**——被封装为 **AI-Ready 数据知识**，作为 Agent 与可视化分析的共享供数层：

- 向 **数据分析 Agent** 供数：驱动自然语言问数、智能报表、归因分析等对话式消费。
- 向 **DashBoard** 供数：作为可视化分析的语义源，避免图表与实际业务口径脱节。

4.5 数据治理 Agent

数据治理 Agent 横向作用于 **多模态数据入湖仓、全模态资产管理、统一语义构建** 三大能力域，通过 Agent 化实现“人工治理 → AI 智能守护”的跃迁。核心功能：

- **质量规则定义 AI 自动治理**：自然语言下发规则，Agent 补齐规则细节并自动挂载至链路。
- **数仓质量诊断**：三类模式（快照、时序、推理），配合 KS、PSI、Wasserstein、卡方、JS 散度等漂移指标。
- **元数据补齐**：Table 与 Column 级描述、业务口径、Owner 的自动补全。
- **资产盘点与敏感字段 AI 识别**：批量识别 PII（Personally Identifiable Information，个人身份信息）、财务、身份等敏感字段，联动数据分级与脱敏，并对全域资产进行盘点。

4.6 统一元数据 TC Catalog

TC Catalog 是整个平台的“元数据中枢”：**Table Catalog、Volume Catalog、Model Catalog、External Catalog** 共享同一套目录、权限、血缘、标签，向上支撑 Agent 与可视化分析，向下打通引擎与存储。

五、引擎与存储层

5.1 计算引擎

引擎	定位	典型场景
批量计算引擎	大规模数据并行处理	批处理 ETL、数仓建仓、模型训练、大规模数据加工
OLAP（Online Analytical Processing，在线分析处理）分析引擎 （规划中）	交互式亚秒响应	交互式分析、报表加速、DashBoard 供数

5.2 湖仓存储 TCLake

- **托管存储 + 读写请求计费**，客户免运维。
- **全模态**：同时承载 Table（结构化）与 Volume（图片、视频、文档、音频、日志）。
- **一份数据，多引擎共享**：批量计算引擎与 OLAP 引擎同源读写，避免数据冗余搬迁。

5.3 客户外部数据源

对客户已有数据源（MySQL、PG 等）通过 External Catalog 与联邦查询能力实现 **免搬迁分析**。

六、架构核心特征

1. **Agent 原生（Agent-Native）**：三大内置 Agent + Harness 体系提供的知识、上下文、记忆、Skill、安全、评测能力，让“对话即交付”成为默认交付形态。
2. **Data + AI 一体化**：数据平台与 AI 平台共用一套元数据、血缘、权限、调度、存储，避免“两套系统”的搬迁与治理成本。
3. **全模态支持**：Table、Volume、Model、External 一体，覆盖结构化、半结构化、非结构化、模型资产、外部镜像。
4. **统一语义驱动**：本体建模与指标、维度体系统一业务口径，是 Agent 与 DashBoard 稳定交付的关键基础。
5. **AI-Ready 数据知识共享**：语义产物封装为供数层，同时服务 Agent 与可视化分析，消除“分析口径漂移”。
6. **一体化治理**：数据治理 Agent 横跨入湖仓、资产管理、语义构建，配合 Harness 的 Agent 安全能力，让 AI 深度介入企业数据仍然可控。
7. **开放连接**：既能托管接入多种外部数据源，也能通过 External Catalog 对客户已有 MySQL、PG 等做免搬迁的联邦分析。

产品功能

最近更新时间：2026-08-12 16:15:00

DataBuddy 覆盖大数据全生命周期，按模块可分为 8 大能力，对应产品文档的 8 个一级目录。

数据接入

智能数据接入：从“人工配置”到“对话式创建”，从“被动运维”到“AI 主动诊断”。

- **40+ 数据源**：关系型数据库、数仓、湖仓、NoSQL、KV、消息队列、流、文件、对象存储、API、多模态
- **5 大接入方案**：
 - 离线同步（批量、增量、定时调度、四大链路）
 - 实时同步（CDC 增量、整库多表、分库分表）
 - 文件到 Table（CSV、Excel、JSON、自动建表）
 - 文件到 Volume（非结构化、图片、视频、文档）
 - 批量到 Volume（FTP、S3、保留目录结构）
- **DataBuddy Agent 智能辅助**：自然语言创建任务、AI 智能问答、任务异常 AI 诊断根因

数据工程

统一数据开发工作台、编排、CI/CD、运维：DIOPS 四大模块闭环，覆盖数据开发全生命周期。

Studio（开发）

- **Notebook**：多语法单元格（Python、SQL、Scala、Markdown）、魔法命令切换
- **SQL IDE**：统一 Spark SQL 方言（Kyuubi）、元数据感知自动补全、参数化查询
- **Python IDE**：跨文件 import、Debug 模式、完整 Python 3 语法
- **智能补全**：Catalog、Schema、Table、Column
- **AI 助手 DataBuddy**：自然语言生成 SQL、代码纠错、自动注释、代码续写
- **Git 原生集成**：commit、push、pull、分支管理

Workflow（编排调度）

- **可视化 DAG**：拖拽添加任务、连线建立依赖、配置触发器、保存即生效
- **9 种任务类型**：Notebook、SQL、Python、Ray、离线接入、IF-ELSE、ForEach、嵌套、质量监控
- **12 种运行条件**：全部成功、全部失败、至少一个成功、至少一个完成 等
- **调度触发**：常规周期（天、小时、分钟、周、月、年）、Cron 表达式、时区
- **持续运行**：前一次终态后 60 秒自动触发，适合近实时场景
- **参数管理**：两层参数、任务间传递、内置参数

CI/CD（发布）

- 配置即代码：基于 YAML、CLI 的 Bundle 发布体系
- 多环境：dev、staging、prod 自动化部署
- 命令集：bundle init、validate、deploy、run、summary、destroy

OPS（运维监控）

- 四层运行监控视图：工作流运行列表、运行记录、运行详情、任务运行详情
- 多渠道告警：Email、Webhook、Teams、企业微信、Slack
- AI 智能诊断：实例信息、根因分析、影响范围、修复建议

数据科学

OneOps 一体化承接企业 AI 工程化：DataOps、MLOps、AIOps。

五大核心模块

模块	能力
Studio	一体化开发 IDE（Notebook、内置 Copilot 辅助）
特征管理 Feature Store	特征 CRUD、内置算子、离在线一致性、跨项目复用
模型实验 Experiment	MLflow 跟踪、AutoML、大模型实验、超参搜索、实验复现
模型管理 Registry	注册、版本管理、模型血缘、冠军模型一键晋级
模型服务 Serving	在线服务、批量预测、多版本 A/B Test、弹性伸缩、自动启停

数据分析

从“等排期”到“对话即洞察”：面向分析师与业务人员。

- 智能问数：自然语言提问，结合语义层秒级取数
- AI 生成分析报告：一键生成完整报告，支持主动定时推送
- 主动智能归因：AI 多维归因，输出可执行建议
- AI 生成看板：自然语言描述，AI 自动规划图表并生成仪表盘
- 数据探索：交互式数据探索、SQL 直查
- 分析空间：创建与管理分析空间、支持分析空间嵌出

数据治理

从“人工治理”到“AI 智能守护”：DataBuddy 四大核心能力。

统一语义层 Unity Semantics

- 本体建模（Ontology）标准化实体、指标、维度、业务逻辑
- 消除口径歧义，避免 SQL 幻觉
- NL2SQL 准确率：无语义层准确率 60% ~ 70%，完整语义层准确率 90%+

数据质量 Data Quality

- 规则引擎：YAML 一站式规则定义、多种规则形态（空值率、唯一性、值域、波动率、产出及时性、自定义 SQL）
- 调度执行：嵌入 ETL 链路、前置阻断、工作流联动
- 规则级精准告警：多渠道推送（企业微信、邮件、Webhook）
- 异常检测：零配置、自学习、Schema 级一键开启
- 数据剖析：三类模式（快照、时序、推理），漂移指标（KS、PSI、Wasserstein、卡方、JS 散度）

多模态数据目录 TCCatalog

- Table Catalog：表、视图、函数
- Volume Catalog：日志、图像、PDF
- Model Catalog：ML 模型、版本管理
- External Catalog：外部数据源镜像（免 ETL 即查）
- 端到端血缘：自动采集 SQL、集成任务，图模式、列表模式双视图

Agent 安全防护

- 双向 Agent Guardrails：提示注入实时检测、阻断攻击链路
- OBO 权限强绑定：真实用户身份，复用底层数据平台权限
- 高危 SQL 拦截、意图校验：调用执行前进行高危判定
- “三要素”精准拦截：不可信输入、状态改变、敏感数据同时命中时，触发人为确认
- 覆盖 OWASP LLM Top10 全部风险

数据应用

Agent、MCP、Model List 构建应用生态。

- Agent 概述：内置三大 Agent 与自定义 Agent 能力
- MCP：Model Context Protocol 原生支持，让人和 Agent 走一套调用链路
- Model List（模型列表）：多模型接入与 Endpoint 管理

Buddy（Agent 能力）

Buddy Agent 是本产品的核心差异化能力，详见 Buddy 一级目录。

- **Buddy 概述：**产品定位与两大场景模式（数据分析、数据开发）
- **Buddy — 数据分析场景：**分析空间管理、复杂数据分析、AI 搭建仪表盘、分析空间嵌出
- **数据工程 Agent：**从“人工搭建”到“Agent 端到端交付”
- **数据治理 Agent：**从“人工治理”到“AI 智能守护”
- **Buddy 数据安全：**Agent 安全防护体系
- **Buddy Agent 基础能力：**知识库、数据复核、@ 引用、Ask 模式、上传文件、产物与产物变更

平台管理

- **工作空间管理：**创建工作空间、成员与权限管理、工作空间设置（通知渠道、Git 连接、个人偏好）
- **数据源管理：**全局数据源配置、添加修改数据源、连通性测试
- **计算资源管理：**计算资源概述、创建、网络设置、用量监控

应用场景

最近更新时间：2026-08-12 16:15:00

DataBuddy 覆盖企业数据全链路核心场景，帮助客户从“人操作工具”进化到“AI 工作、人把关”。以下是三大典型场景。

数据工程—从“人工搭建”到“Agent 端到端交付”

场景介绍

数据工程师承接业务方的数据需求（如“每天早上 10 点看到销售绩效、季节性趋势与商品品类表现”），需要完成数据源接入、ETL 代码开发、工作流编排、调度配置、告警设置等全链路工作。传统模式下：**需求理解、资产探查、方案设计全靠人工，耗时费力。**

典型客户：互联网、零售、金融行业的数据平台部门、数据仓库团队；面向业务方（运营、市场、财务等）持续交付数据需求。

痛点与方案对比

环节	传统模式痛点	DataBuddy 方案
需求理解与方案设计	需求理解、资产探查、方案设计全靠人工	AI 理解需求生成方案： 理解业务需求、探查现有资产、自动生成数仓设计方案
编码与工作流配置	建表、ETL SQL、脚本开发繁琐枯燥，工作流配置易漏配	AI 编码、配置工作流： ETL 脚本、Pipeline 一键生成，自动配置任务依赖、调度、告警、重试策略
运维与根因分析	运维凭经验，手动翻日志、排查上游任务，链路长、根因分析依赖个人经验	智能运维： 自动解析日志、代码、资源、上游依赖，分析失败原因，推荐可信修复方案
规范治理	数仓分层、代码规范、命名规则等规范依赖人工执行，易遗漏或不一致	规范落地为 Rules： 数仓分层策略、代码规范、命名规则、告警策略等定义到 Rules，工程环节自动应用

量化收益

- 重复开发工作量降低 80%
- 人效提升 5 到 10 倍
- 端到端完成典型数据需求：数据源接入、ETL 代码、工作流 DAG、沉淀到 Catalog 一条链路走通

数据治理—从“人工治理”到“AI 智能守护”

场景介绍

数据治理团队负责保障数仓质量、敏感数据识别、根因诊断、全仓巡检。传统模式下重人力投入，且难以覆盖存量大批量的关键表。

典型客户：金融、政企、大型互联网的数据管理团队与数据安全团队；面对合规审计、数据资产盘点、跨部门数据质量管控等常态化诉求。

痛点与方案对比

环节	传统模式痛点	DataBuddy 方案
敏感数据识别	敏感数据靠人工标记，漏配置或规则不全都会导致敏感数据“裸奔”	AI 语义识别，自动打标分级： AI 理解数据画像与字段语义，自动识别敏感数据
质量规则设计	质量规则凭经验拍脑袋，关键规则遗漏、冗余堆积，存量关键表无质量监控	AI 驱动，推荐规则： 理解字段语义，自动推荐完整性、唯一性、一致性等规则，一键审核
根因诊断	问题诊断止于“表象”：不知道是上游业务逻辑调整还是数据任务故障，结论停在“表 X 有问题”	AI 智能溯源，根因诊断： 获取上下游血缘、任务状态、资源状态，直达根因
全仓巡检	人工巡检覆盖率低，人工写 SQL 检查数仓，问题发现周期长、修复慢	AI 全仓扫描、自主守护： AI 自主多维扫描，及时发现问题、诊断根因、执行修复

量化收益

- 人工治理工作量降低 80%
- 人效提升 5 到 10 倍
- 7×24 主动守护：AI 常态化扫描存量与增量资产，问题第一时间发现并触达

数据分析—从“等排期”到“对话即洞察”

场景介绍

业务方（销售、运营、财务、决策者）需要及时的数据洞察辅助决策，但传统模式下依赖 IT 排期，数据触达链条长、时效差。

典型客户：零售、电商、金融、互联网中的业务部门（销售、运营、市场、财务、决策层）；日常需要拉数、看板、异动归因等自助分析能力。

痛点与方案对比

环节	传统模式痛点	DataBuddy 方案
取数排期	业务提需求，等待 1 到 3 天	智能问数： 自然语言提问，结合语义层秒级取数

报告制作	报告人工制作：人工构思、取数、做图，天级交付	一键报告 & 主动推送： AI 生成完整分析报告，支持主动定时推送
异动归因	异动归因靠经验逐层下钻，耗时费力	主动智能归因： AI 多维归因，输出可执行建议
看板开发	看板开发跨部门提需求，周级交付	AI 生成看板： 自然语言描述，AI 自动规划图表并生成仪表盘

量化收益

- 分析耗时降低 80%
- 分析效率提升 5 到 10 倍
- 分钟级分析速度：从提需求到看到结果压缩到分钟级

典型客户角色与协同

一条 DataBuddy 流水线可贯穿五类角色：

角色	关注点
业务负责人	把控 AI 方案的业务价值
数据工程师	构建数据工作流
数据科学家	业务问题转算法、训练调优
研发工程师	部署 AI 模型到生产环境
数据管理员	保证数据管理与合规性