

大数据智能体工作台 DataBuddy Buddy



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

Buddy

Buddy — 数据分析场景

创建与管理分析空间

数据工程 Agent

Buddy

Buddy — 数据分析场景

创建与管理分析空间

最近更新时间：2026-08-12 16:15:00

概述

分析空间是 Buddy 数据分析模式的核心上下文，把表、指标模型、知识库聚合到一个业务相关的小范围内。本文介绍如何创建分析空间、向空间添加资产、管理空间内数据，以及维护空间。

前提条件

- 已开通 DataBuddy 服务并完成空间初始化。
- 已切换到 Buddy Agent 的 **数据分析** 模式。
- 已在 Catalog 中准备好数据表，或在 **语义模型** 中已发布的指标模型。

使用限制

限制项	说明
模式	仅在“数据分析”模式下可见与可创建
数据来源	仅可纳入已在 Catalog 中可见的表，以及已在指标平台发布的指标模型
创建归属	空间由个人账号创建，可分享给团队成员，分享后由您作为管理员维护

操作步骤

步骤 1：进入分析空间管理入口

1. 登录 [DataBuddy 控制台](#)。
2. 在工作空间列表中，单击目标工作空间进入。
3. 左侧导航单击 **Buddy Agent**，在模式切换器中选择 **数据分析**。
4. 在左侧“分析空间”区域，单击**空间切换器**。

切换器支持按名称搜索现有空间，列表底部提供 + **创建分析空间** 入口。

步骤 2：创建分析空间

1. 在空间切换器中单击 + **创建分析空间**。

2. 在创建页选择 **指标模型** 与 **数据表**（与数据表至少二选一），可勾选指标平台已发布的语义模型，以及 Catalog Schema 下的表。每个空间内最多添加 30 个数据表和指标模型。
3. 填写空间名称（必填）与描述（选填）。
4. 单击**确定**。

提示

首次使用时如果当前 Workspace 没有任何分析空间，Buddy Agent 会预置一个默认分析空间引导您快速体验。

步骤 3：管理空间内的数据

进入空间后单击左上角 **数据** Tab，进入空间数据管理页。顶部支持按 **指标模型** / **数据表** 类型过滤，单击任一记录打开详情：

- **数据表详情**：表描述、字段列表（字段类型 / 字段名 / 字段描述 / 最多 5 个示例值）。
- **指标模型详情**：字段分为“指标”与“维度”两类分别展示。

提示

表详情与指标模型详情只读。元数据统一由 **数据目录（Catalog）** 与 **语义模型** 管理，如需修改字段描述、添加示例值，请到对应页面编辑。

步骤 4：在空间内提问

确认数据已纳入空间后，可在对话区直接提问。Buddy Agent 会只在当前空间内的表、指标模型、空间知识库中检索。

步骤 5：通过更多菜单管理空间

在**空间切换器**中，将鼠标悬停在目标分析空间上，单击右侧的 **更多菜单（…）**，可执行以下操作：

操作	说明
空间信息	修改空间名称、描述、示例问题
收藏空间	收藏后的空间在列表中排在最上方
分享空间	添加空间成员，分享后由您作为管理员维护
删除空间	二次确认后删除，空间内的会话与产物列表一并清理

权限模型

分析空间内角色与可执行操作：

操作	管理员	编辑者	使用者
使用空间对话、查看资产、查看知识库	✓	✓	✓
增删改空间内表 / 指标模型 / 知识库	✓	✓	✗
修改空间内成员	✓	✗	✗
删除空间	✓	✗	✗

⚠ 注意

分享空间不会赋予被分享者访问 Catalog 中表的权限。被分享者只能基于他自己在 Catalog / 指标平台拥有的权限读取空间内资产；无访问权限时提问触达这张表会失败。空间成员关系与底层资产权限是两套独立体系，详情请参见 [成员与权限管理](#)。

常见问题

为什么我在 Catalog 中能看到的某张表，在创建分析空间时却选不到？

通常是以下两个原因：

- 您的账号对该表只有 Catalog 列表查看权限，没有内容读取权限。请联系 Catalog 管理员授予权限。
- 该表所在的 Catalog 不是 Workspace 默认 Catalog 且未挂接到您的 Workspace。

分享给同事的空间，对方为什么看不到？

请检查：

1. 是否使用“更多 > 分享空间”入口添加对方账号。
2. 对方是否切换到了“数据分析”模式。
3. 对方是否对空间内资产有访问权限。

可以把一张表纳入多个分析空间吗？

可以。表与指标模型在 Catalog / 指标平台中是统一存储的，分析空间只是“引用”这些资产，不会复制数据。

数据工程 Agent

最近更新时间：2026-08-12 16:15:00

数据工程 Agent 是 Buddy 在“数据开发”模式下的 AI 数字员工，您只需用自然语言描述需求，Agent 即可自动完成从需求理解到任务上线的完整数据工程链路。

AI 驱动的全流程

传统数据工程需要工程师手动梳理需求、逐行编码、逐个配置任务、手动排查问题，整个过程耗时费力、响应慢、易出错。数据工程 Agent 将这一切变为 AI 驱动的自动化流程：

传统方式	Buddy 数据工程 Agent
人工梳理需求	AI 理解需求
手动编码	生成数仓方案
逐个配置任务	自动生成代码
手动运维排查	智能编排 workflow
耗时费力、易出错、效率低	高效智能、响应快、准确率高

数据工程 Agent 覆盖 6 大核心能力，形成完整闭环：

步骤	能力	说明
1	需求生成数仓方案	理解业务需求，自动生成贴合业务的数仓建设方案
2	智能数据接入	自动识别数据源、智能映射字段，快速完成数据接入
3	代码生成	根据方案与规则自动生成高质量 SQL / 脚本代码
4	工作流自动开发	自动编排任务依赖与调度，一键生成完整 workflow
5	数仓结构分析	自动解析现有数仓结构，识别分层关系与建模问题，给出优化建议
6	任务诊断修复	实时监控任务，智能诊断问题并提供修复建议

整个流程由 AI 端到端串联，您只需在关键节点确认即可，无需在多个模块间手动切换操作。

与数据分析 Agent 的差异

Buddy 通过顶部「场景模式」切换在 数据开发模式（数据工程 Agent）和 数据分析模式 之间切换。两个模式产出的工件、知识库、产物范围互相隔离：

维度	数据工程 Agent（本文档）	数据分析 Agent
主要读者	数据工程师 / 数据开发	数据分析师 / 业务方
典型产物	SQL 文件、Python 文件、Notebook、 工作流、数据接入任务	Markdown 报告、HTML 仪表盘、查询 结果表、图表
知识库归属	工作空间级知识库	分析空间级知识库（与该分析空间绑定）

前提条件

在使用数据工程 Agent 前，请确保：

- 工作空间已开通 Buddy。
- 当前账号对目标 Workspace 内涉及操作的资源（数据源、Catalog、Studio 文件夹、计算资源、工作流）拥有相应权限。Agent 在执行落地操作时会校验该账号的真实权限，无权限时会给出权限不足提示。

场景一：智能数据接入

功能说明

智能数据接入支持通过自然语言描述创建实时 / 离线同步任务，覆盖以下场景：

- **实时整库多表接入**：把整个库（或正则匹配的表集合）实时同步到 Catalog。
- **实时分库分表接入**：把多个分库的同名表合并写入一张目标表。
- **离线单表接入**：每天 / 每小时定时同步单张源表到目标表。
- **离线批量单表接入**：一次性为多张源表批量创建离线同步任务。

Buddy 自动识别接入方式、生成字段映射、推荐写入模式与调度时间，并在关键节点等待您确认后落地。您只需提供**数据源名称、源库/表范围、目标 Catalog**，其余配置（任务名、同步策略、增量字段、写入模式、字段映射等）由 AI 智能推荐。

使用方式

在 Buddy 输入框用自然语言描述即可，您只需提供**数据源名称、源库/表范围、目标 Catalog**：

输入示例	效果
「把生产环境 MySQL 的 orders 表每天同步到 lakehouse_catalog」	创建离线单表同步任务
「mysql_prod_01 里所有以 order_ 开头的表实时同步到 ods_catalog」	创建实时整库多表同步任务
「order_db_00 到 order_db_07 这 8 个分库的 order 表合并同步到 order_all」	创建实时分库分表同步任务

「把 ods schema 下的所有表每天同步到 lakehouse_catalog」

创建离线批量单表同步任务

场景二：代码辅助

功能说明

代码辅助支持在 Studio 内通过自然语言完成代码全生命周期操作：

- **代码生成**：自然语言描述生成 SQL / Python / Notebook 文件，自动插入 Studio。
- **代码解释**：对当前代码进行自然语言解析，说明核心逻辑与业务含义。
- **代码纠错与方言转换**：自动语法纠错，转换 `:param` ↔ `${param}`、`dbutils` → `dlcutils` 等方言差异。
- **试运行 + 错误自动修复**：脚本编写后自动试运行，运行失败时分析错误日志并尝试修复重试。

Buddy 与 Studio 文件树、编辑器、运行结果面板深度联动，修改文件时展示 Diff，您可以选择接受或拒绝。

使用方式

在 Buddy 或 Studio 侧边栏中用自然语言描述即可：

输入示例	效果
「帮我写一个 SQL，从 ods_orders 统计每天每个渠道的 GMV」	生成 SQL 文件
「新建一个 Python 文件，用 dlcutils 读取 dws_user_daily 表」	生成 Python 文件
「解释这段代码做了什么」	代码解释
「把这段 Databricks SQL 改成 DataBuddy 方言」	方言转换
「运行这个 SQL」	试运行，失败时自动修复重试

场景三：工作流编排

功能说明

工作流编排支持通过自然语言把 Studio 中的脚本组装成工作流：

- **从代码创建任务**：基于 SQL / Python / Notebook 文件创建任务节点。
- **DAG 自动生成**：解析代码中的 `FROM` / `JOIN` 关系，自动生成任务依赖与 DAG 拓扑。
- **DAG 校验**：循环依赖检测、不合理依赖识别、调度冲突检测。
- **调度配置**：根据业务时效推荐调度周期、调度时间、依赖策略、重试策略。

- **workflow增删改查**：自然语言修改节点、配置、依赖。

使用方式

在 Buddy 输入框用自然语言描述编排需求即可：

输入示例	效果
「把 dwd_orders.sql、dws_order_daily.sql、ads_channel_gmv.sql 串成一个工作流，每天 03:00 调度」	创建工作流 + 配置调度
「在 ads_channel_gmv 后面加一个推送任务」	新增节点 + 配置依赖
「把整个工作流的调度改成每小时一次」	修改调度周期

场景四：智能诊断

功能说明

智能诊断覆盖数据工程全链路的异常定位与修复建议，覆盖以下范围：

- **workflow任务诊断**：工作流中任务出现 **失败 / 等待资源 / 上游失败** 时，自动根因分析、影响范围评估、失败链路追踪，并提供修复建议（重跑 / 终止 / 修改后重跑）。
- **Studio 内代码运行诊断**：在 Studio 中运行 SQL / Python / Notebook 报错时，分析错误日志并给出修复建议。
- **实时接入任务诊断**：实时数据同步任务出现延迟、断流、字段异常等问题时，定位根因并给出处理建议。
- **离线接入任务诊断**：离线同步任务失败、超时、DDL 变更等情况时，定位根因并给出处理建议。

使用方式

在工作流运行详情中 hover 失败状态节点，点击 **AI 诊断** 即可触发；也可以在 Studio 运行面板中点击 **AI 诊断**，或在 Buddy 中用自然语言描述：

输入示例	效果
「xx 工作流任务失败了，帮我看看什么原因」	根因分析 + 影响范围评估
「xx 任务的 SQL 跑失败了，帮我看下」	Studio 内代码运行诊断
「xx 实时同步任务延迟严重，帮我排查」	实时接入任务诊断
「xx 离线同步任务报错了，帮我查下原因」	离线接入任务诊断

端到端场景示例：从 0 开始新建数仓

以下通过一个完整的场景展示数据工程 Agent 如何从一句需求出发，端到端完成数仓建设全流程。

输入示例

我是一个数据工程师，现在业务团队新增了一个外部 PostgreSQL 数据源，希望基于这批数据建设一套业务分析数仓。

业务人员希望每天查看销售绩效、季节性销售趋势和商品品类分析，包括 GMV、订单量、收入、客单价、品类销售占比和热销品类排行等指标。

数据源名称为 XX，源表位于 XX schema 下。我希望将该 schema 下的所有源表同步到 XX catalog 中，并基于这些源表完成数仓设计与建设。数据每天早上 10 点更新。

Agent 自动执行流程

