弹性 MapReduce

AI 探索中心



腾讯云

版权所有：腾讯云计算（北京）有限责任公司

【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

AI 探索中心
操作指南
快速开始
大数据专家团
数据源与知识库
代码工作台
API 调用
EMR 连接器与 Skill 最佳实践

AI 探索中心
操作指南
快速开始

最近更新时间：2026-11-20 11:22:03

简介

AI 探索中心 提供面向企业数据自主分析与洞察的智能体服务。您可通过本文档快速体验产品功能。

前置条件

您已在腾讯云国内站 注册账号 并完成实名认证。

开通使用

登录 EMR 控制台，在一级菜单 Data+AI 中，单击 **AI 探索中心**。阅读《腾讯云服务协议》，**单击同意授权并开通**。同意授权后，您可开始使用 AI 探索中心。

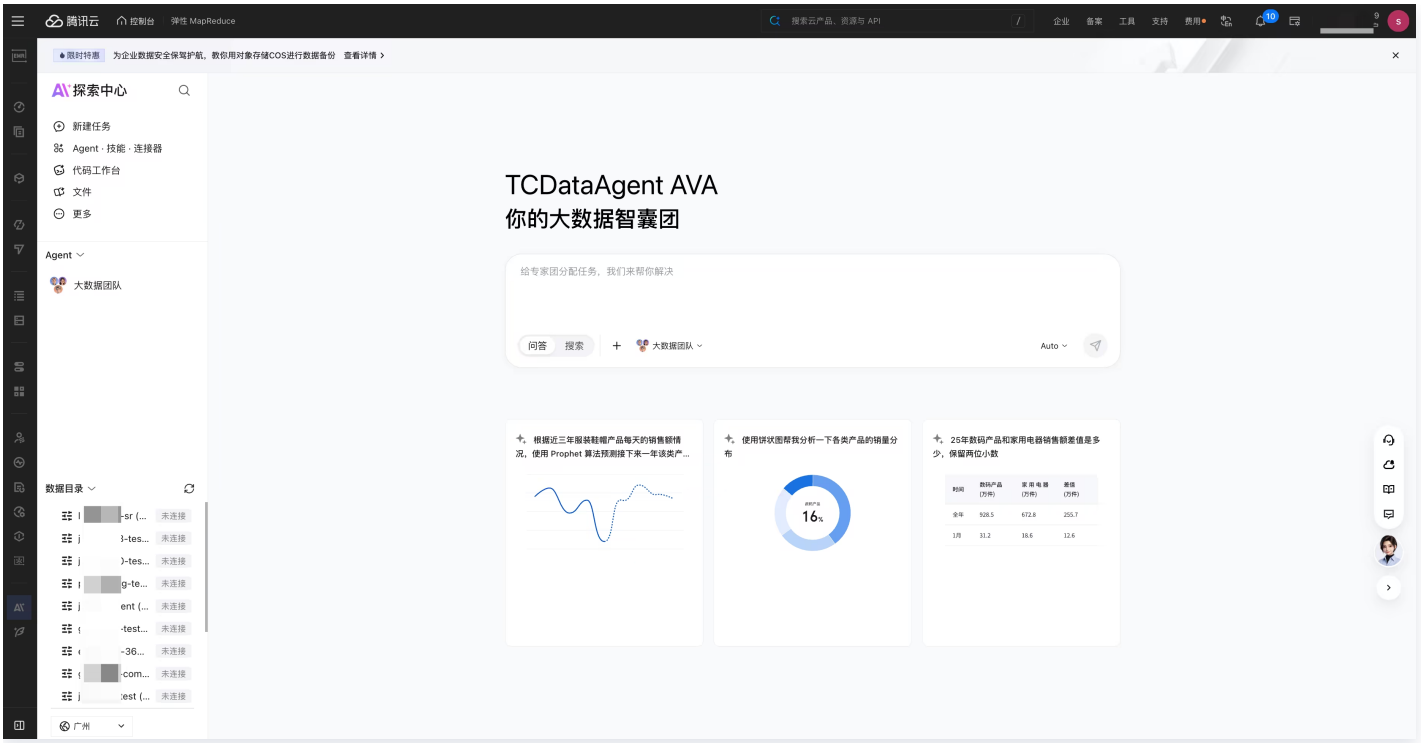
使用样例数据进行问答

AI 探索中心内置了官方样例数据和问题。您可以直接在 AI 探索中心首页，点击示例问题（如：用 Prophet 算法预测服装鞋帽年销售额）进行体验。
官方数据中包含 产品表（products）、订单表（orders）、order_items、users 表，您可以基于官方数据源进行自由提问，选择部分或全部使用默认设置。
提问后，AI 探索中心首先会对用户提问进行解析，识别用户意图并规划子任务。

在“用 Prophet 算法预测服装鞋帽年销售额”示例中：

- 第一个子任务是加载服装鞋帽销售数据，Agent 将调用 SQL 工具加载数据，此时自动进行 NL2SQL、运行 SQL 获取数据等操作，按日期分组计算每天的销售额总和。
- 第二个子任务是使用 Prophet 模型训练与预测，Agent 使用 Prophet 算法对服装鞋帽产品的销售额进行时间序列预测，最后生成可视化图表展示预测结果和置信区间。
- 在完成上述子任务后，Agent 还将根据历史数据与模型预测结果，给出洞察与建议，总结核心发现与结论，提供优化方向与行动计划建议。

版权所有：腾讯云计算（北京）有限责任公司


腾讯云
控制台
弹性 MapReduce
搜索云产品、资源与 API
企业
备案
工具
支持
费用
限时特惠
为企业数据安全保驾护航，数你用对象存储COS进行数据备份
查看详情
探索中心
新建任务
Agent · 技能 · 连接器
代码工作台
文件
更多
Agent
大数据团队
数据目录
广州
TCDataAgent AVA
你的大数据智囊团
给专家团分配任务，我们来帮你解决
问答
搜索
大数据团队
Auto
根据近三年服装鞋帽产品每天的销售额情况，使用 Prophet 算法预测接下来一年该类产...
使用饼状图帮我分析一下各类产品的销量分布
16%
25年数码产品和家用电器销售额差值是多少，保留两位小数
时间
数码产品（万件）
家用电器（万件）
差值（万件）
全年
928.5
672.8
255.7
1月
31.2
18.6
12.6

大数据专家团

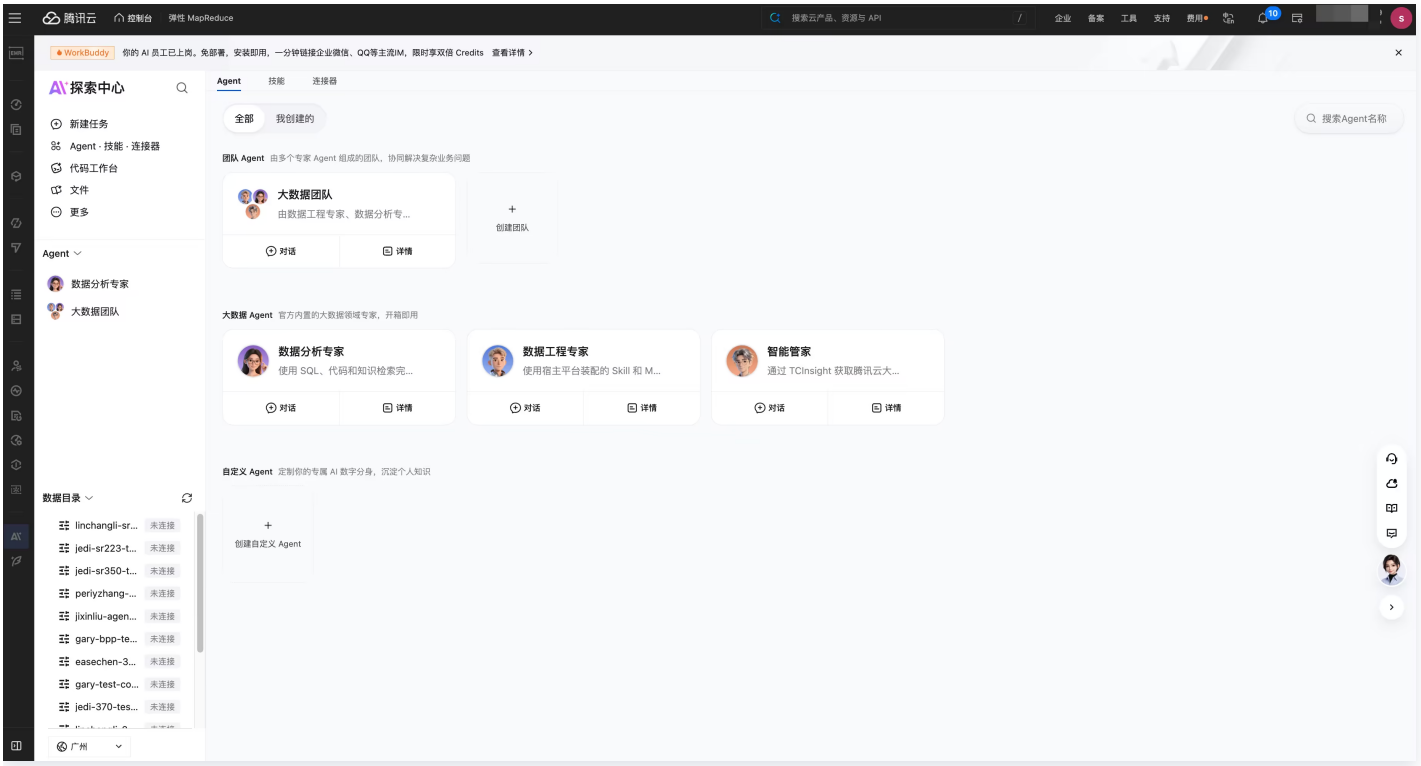
最近更新时间：2026-11-20 11:22:03

大数据平台的日常运营涉及数据分析、集群运维、任务调度、资源管理、安全审计等多个专业领域。当前团队面临以下核心痛点：

- 工具碎片化：EMR 控制台、HDFS 管理、Spark 监控、告警系统……运维和分析人员每天在 5-8 个系统间切换；
- 知识孤岛：老员工的运维经验停留在个人笔记里，新人上手周期长达 2-3 个月；
- 协作断裂：一个"华东区用户活跃度分析"任务涉及数据抽取、计算、可视化多个环节，需要 3-4 人接力完成，沟通成本高；
- 重复劳动：常见的巡检、报表、故障排查流程高度标准化，但仍依赖人工逐步操作。

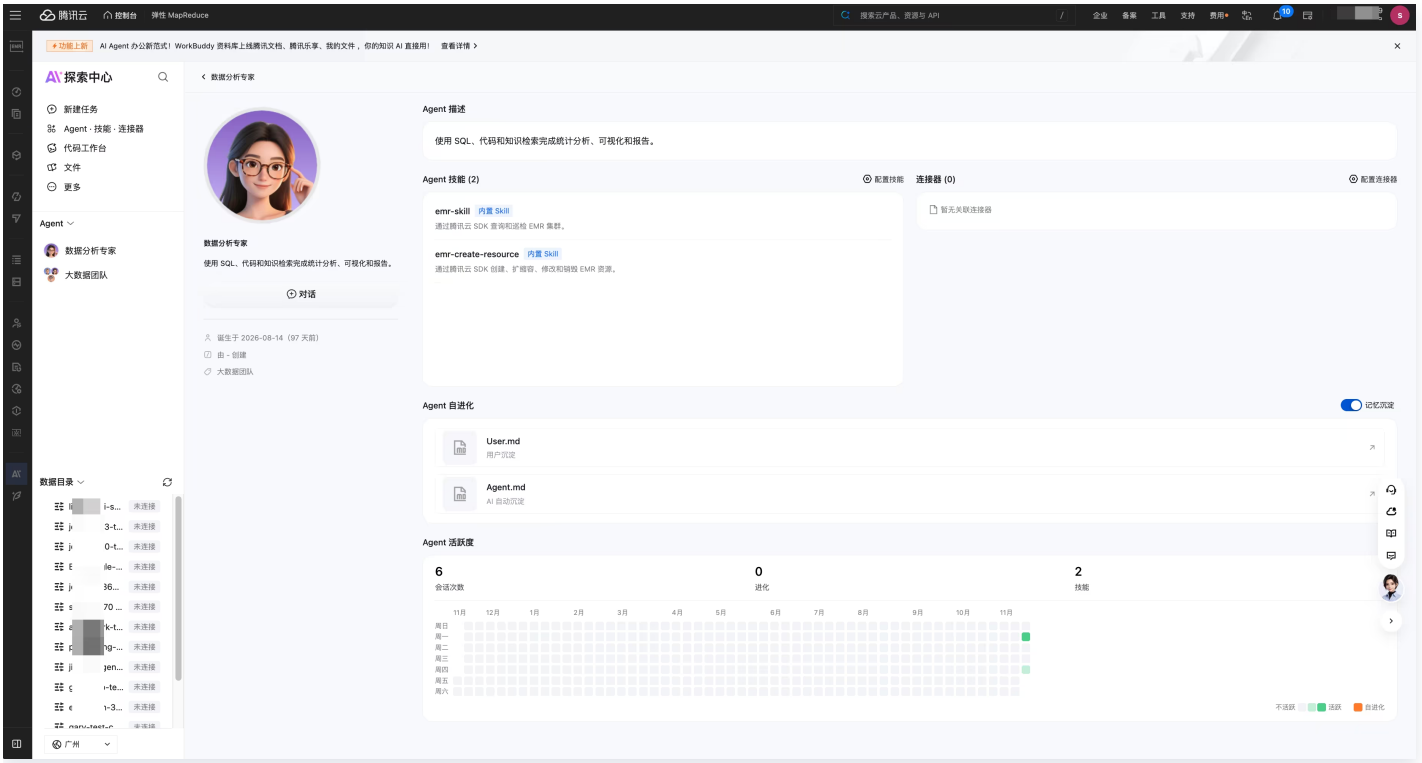
AI 探索中心预置的专业大数据专家与专家团，开箱即用。当前版本内置三个大数据领域专家与一个大数据专家团，并支持自定义 Agent 创建：

- **数据工程专家：**负责多模态数据处理、数据质量管控、ETL、任务开发，构建高效可靠的数据加工链路，将原始数据转化为可分析的高质量资产。
- **数据分析专家：**负责从海量数据中提取关键洞察，构建数据模型与可视化报告，为业务决策提供数据驱动支持。
- **智能管家：**负责集群监控、性能调优、故障排查与容量规划，确保数据平台高可用。



AI 探索中心支持为 Agent 配置相关技能、连接器。基于 Hermes Agent 实现自进化，在解决复杂任务或排查问题后，将有效路径提炼为新的技能或更新已有规则，避免同类问题重复摸索，提高解决问题效率。同时平台记录

Agent 活跃度，包括会话次数、进化次数、配置技能。

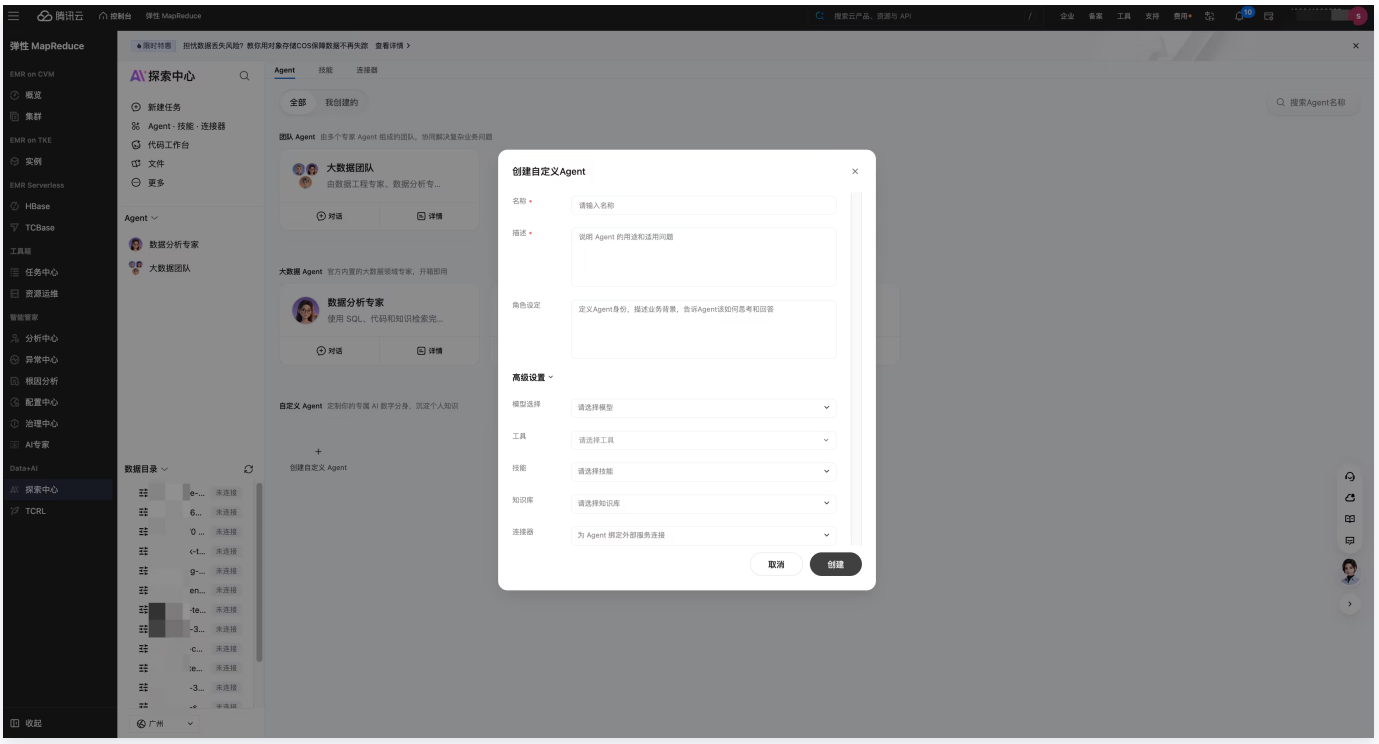


快速组建自定义 Agent 与团队

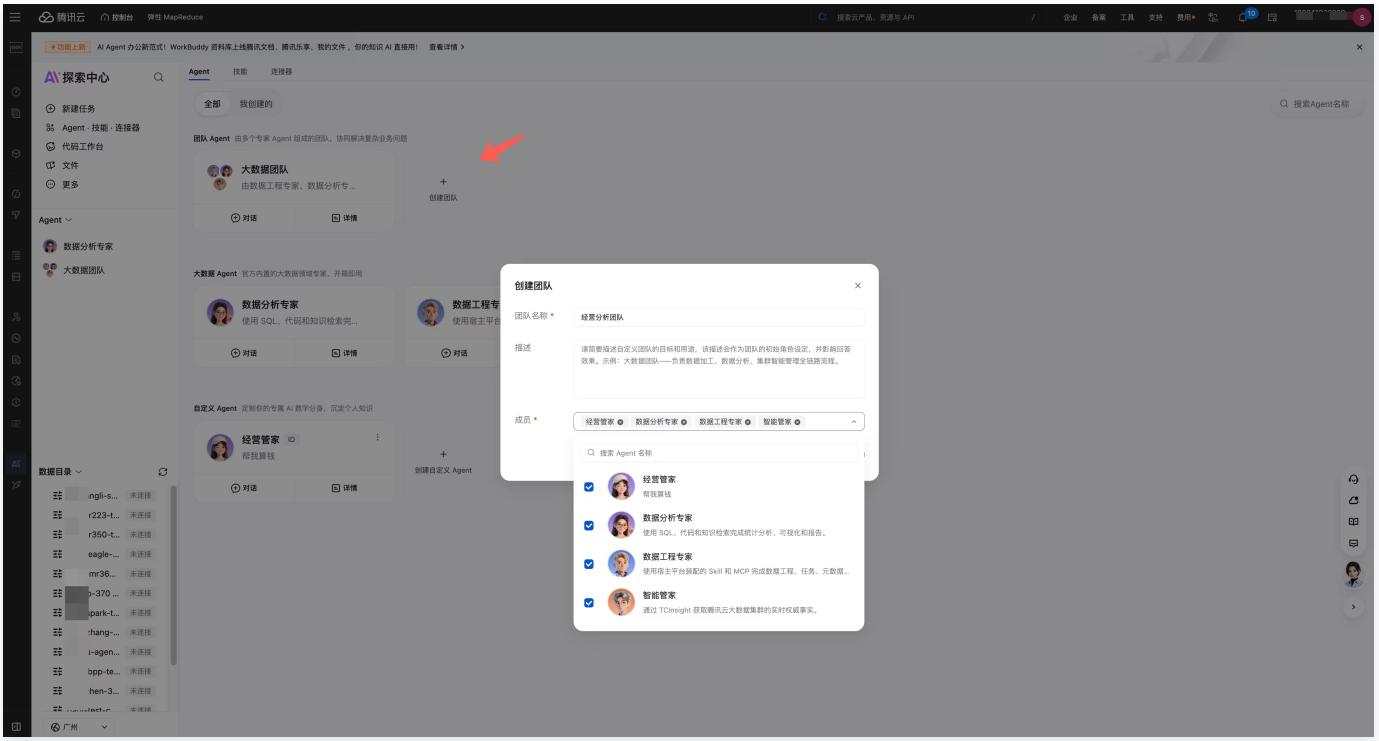
除开箱即用的内置专家与大数据专家团外，AI 探索中心支持结合企业自有数据规范与业务分工，快速创建自定义 Agent 并组建成专属的协作团队，将跨环节的复杂任务交由分工明确的智能体协同完成：

- 明确职责与创建自定义 Agent：围绕特定业务场景或细分岗位（如湖仓成本治理、实时作业排查、业务指标口径审计等）创建自定义 Agent，设定角色定位、关联知识库，沉淀团队专属的领域知识与操作 SOP。
- 绑定技能与连接器：为自定义 Agent 配置所需技能（Skill）与数据源/工具连接器（如 EMR、数据目录、SQL/计算引擎或 MCP 服务），赋予其实际检索、分析与执行能力；同时开启记忆沉淀与自进化机制，让 Agent 在持续执行与排障中不断积累有效路径。

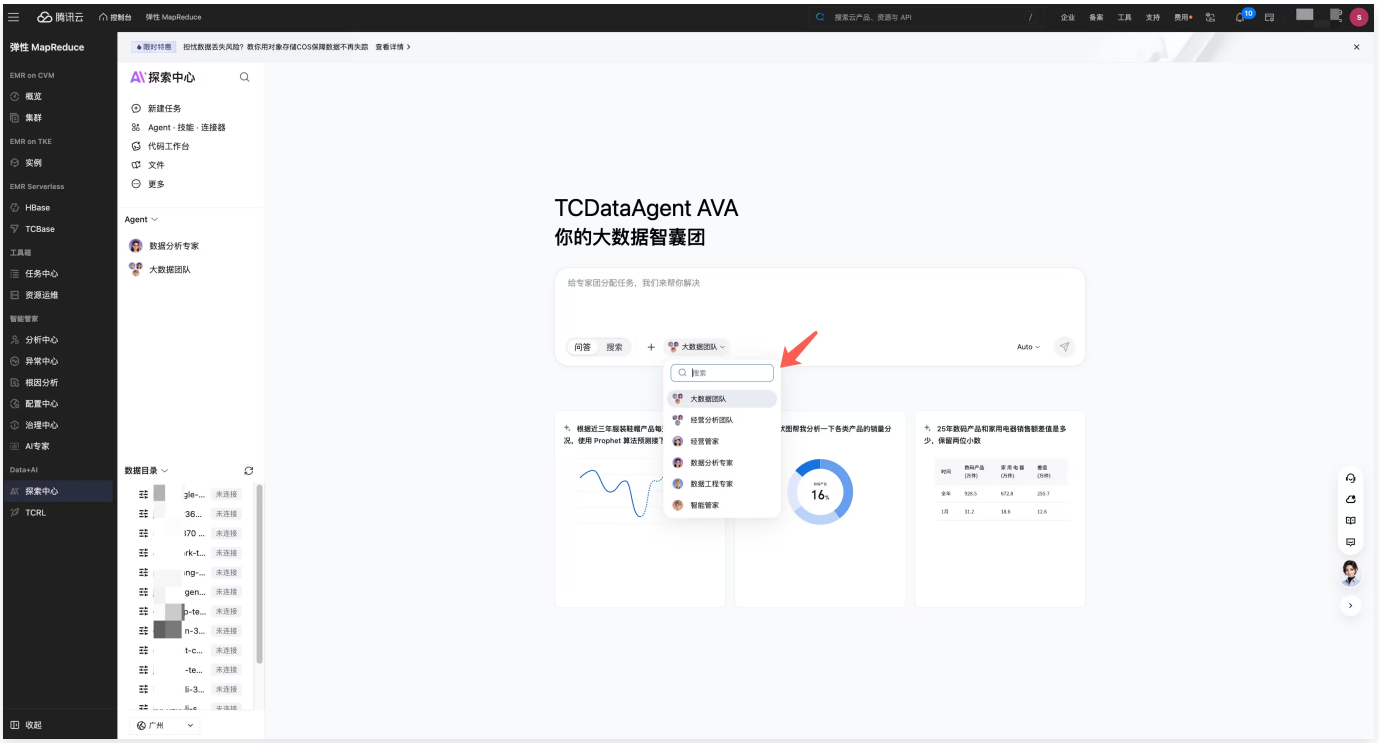
版权所有：腾讯云计算（北京）有限责任公司



- 组建协同团队并投入使用：将具备不同专长的内置专家与自定义 Agent 编排进同一团队，按任务链路分工协作（例如由数据工程专家完成数据准备与校验、数据分析专家负责建模与洞察输出、智能管家辅助集群与任务保障），减少多系统切换与人工接力成本，实现端到端的自动化分析与运维闭环。



- 在通过自然语言对话任务时，可以根据需要选择自建团队或 Agent。


弹性 MapReduce
探索中心
新建任务
Agent · 技能 · 连接器
代码工作台
文件
更多
Agent
数据分析专家
大数据团队
数据目录
TCDataAgent AVA
你的大数据智囊团
给专家团分配任务，我们来帮你解决
问答
搜索
大数据团队
Auto
大数据团队
经营分析团队
经营管家
数据分析专家
数据工程专家
智能管家

数据源与知识库

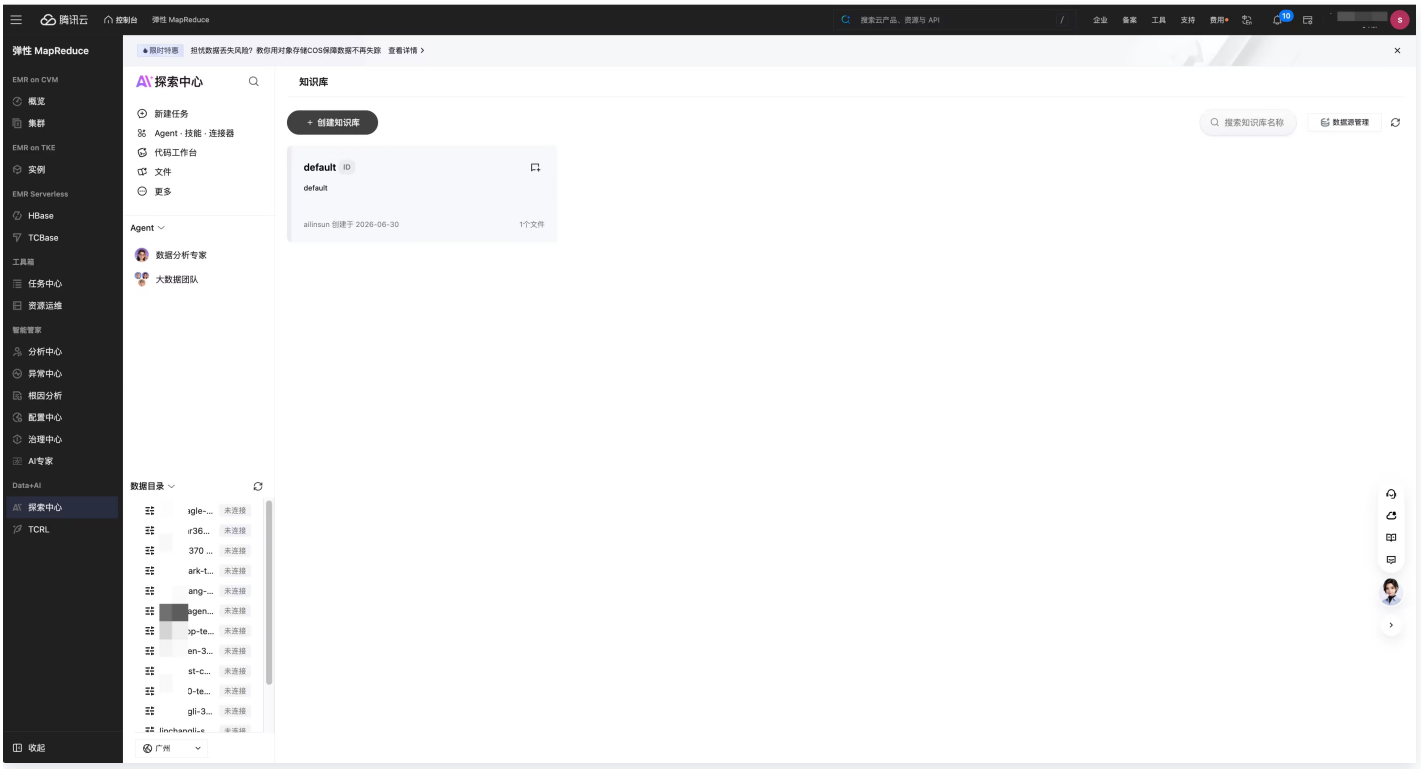
最近更新时间：2026-11-20 11:22:03

操作场景

AI 探索中心支持用户自建数据知识库，支持将各类文本文档及数据库表解析为可问答的知识资产。自建数据知识库后，您可以在首页勾选相应的知识库，AI 探索中心将提取文本内容、语义信息，根据知识库内容与您进行对话问答，分析挖掘知识库的数据价值并提供更专业的回答。

操作步骤

1. 在 AI 探索中心，单击**更多**，选择 **数据知识库** 功能，单击**创建知识库**按钮，输入知识库名称与描述。
2. 单击**添加文件**，根据需要选择添加文件与库表方式。目前支持四种方式：从本地上传、从对象存储 COS 导入文档、从数据源关联数据表、从网页链接添加。
3. 单击**开启新对话**，勾选创建的知识库，进行智能问答。



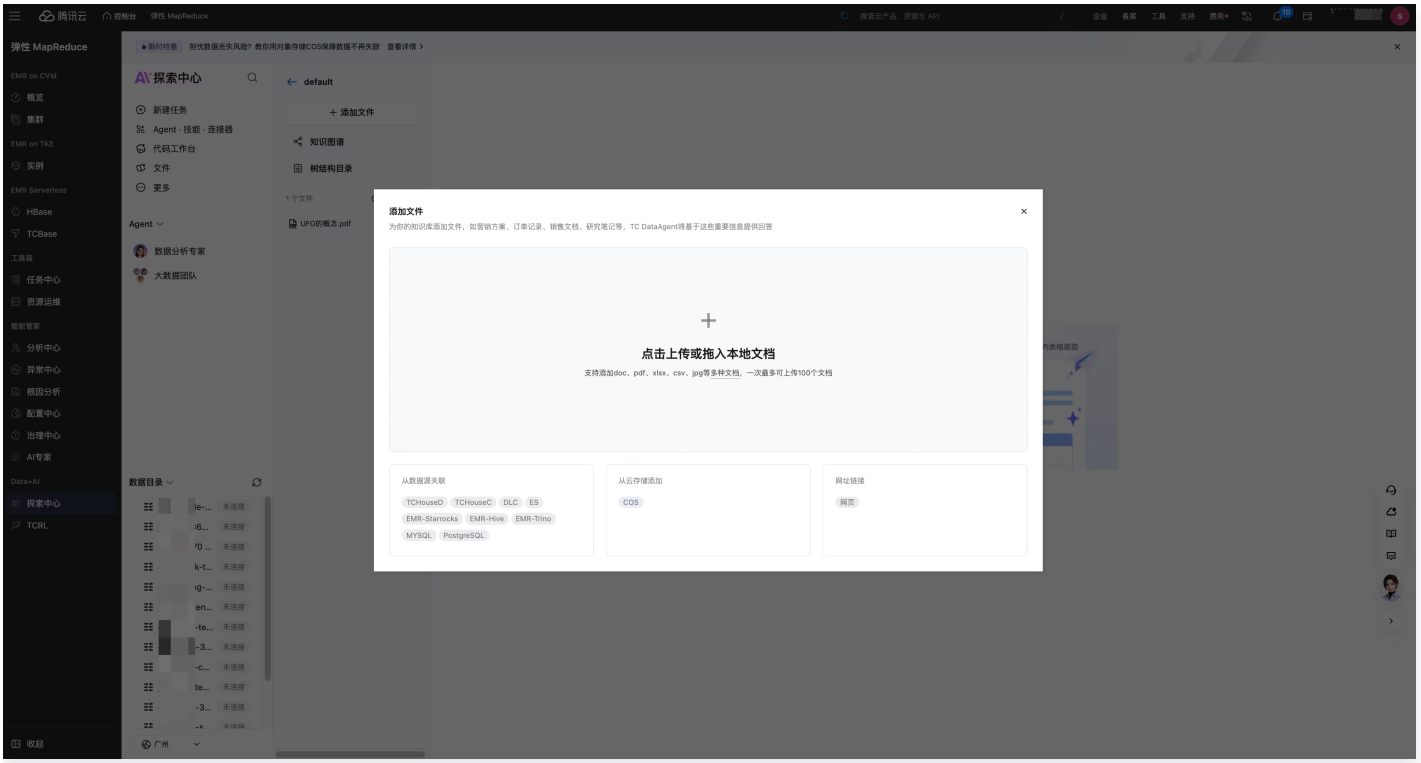
添加文件

向知识库添加文档和数据时，您可以选择以下三种方式：

方式一：从本地上传文档：

1. 进入已经创建的知识库，单击**添加文件** > **从本地上传文档**，您可单击上传本地文档，或将本地文档拖拽至目标区域。
2. 上传后，您可进行“原文件预览”与“文本分段”预览。也可以继续**点击上传或拖入本地文档**，一次最多可上传100个文档。
3. 在文档分段方式上，AI 探索中心默认使用根据语义自动分段的**智能分段**方式。若您需自行设置分段规则，可以在文档导入后进行单击**调整配置**进行修改。调整配置说明如下：
   - **精准解析**：默认选择，支持解析图片和表格中的文本
   - **知识增强**：支持文档概要和段落概要总结，开启后有助于提升检索的召回率
   - **分段策略**：默认使用智能分段，如有特殊需求，可调整为自定义分段，自定义分段配置内容如下：
4. 配置完成后，单击**保存并重新处理**，文档会重新进入解析中状态。

字段名称	字段说明	字段值说明
分段标识符	用于识别文本中自然分段边界的字符或字符串	可选值：换行符“\n”、中文逗号“，”、中文句号“。”、中文叹号“！”、中文问号“？”、英文逗号“,”、英文句号“.”、英文叹号“!”、英文问号“?”、自定义
分段最大长度	单个文本块所允许包含的最大文本字符数	最小值500，最大值5000，步长1
分段重叠度	相邻两个文本块之间重复的文本比例	最小值0%，最大值100%，步长1%



方式二：从对象存储 COS 导入文档

1. 进入已创建的知识库，单击**添加文件 > 从 COS 导入文档**。您可在弹窗中选择目标 COS 路径和文件，单击**确定**。
2. 进入文档配置页面。您可选择继续**从 COS 导入文档**，后续操作细节同上文 从本地上传文档 。
3. **重命名与删除文档：**在文档上传后，点击**文档列表**中“···”，可进行文档重命名或删除操作。

当前支持的文档类型说明如下：

- 文本类：PDF、DOC、DOCX、PPT、PPTX 100M 以内，MD、TXT 10M 以内，PNG、JPG、JPEG 20M 以内
- 表格文档：XLS、XLSX、CSV 100M 以内，文件最多支持一个 sheet 工作表，自动识别表格第一行作为表头。

> **说明：**
> 单次最多可上传或导入100个文档。
> 从 COS 导入文档时，暂仅支持选择腾讯云北京地域的对象存储 COS 桶。
> 对于图片文件，暂仅支持提取图片中的文本内容，更多图片语义理解功能敬请期待。

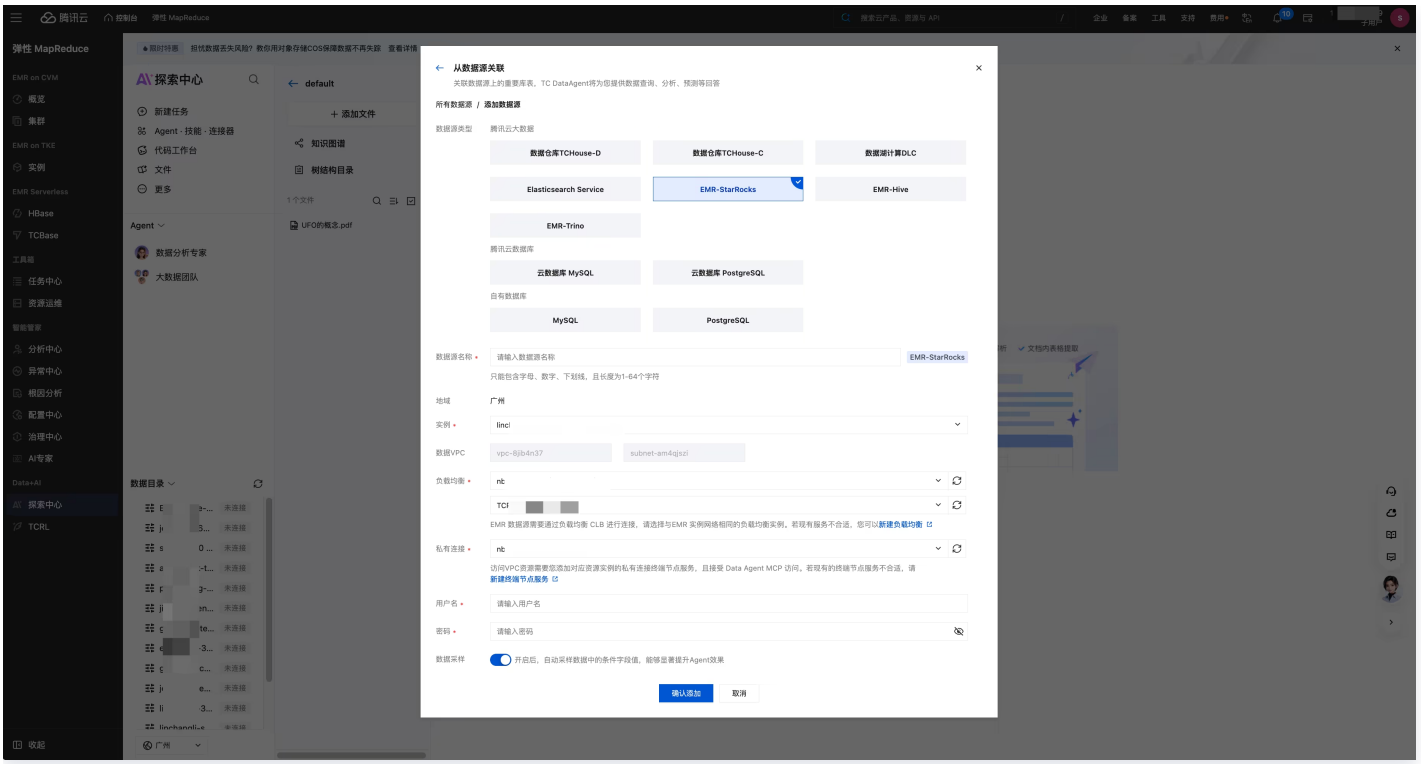
方式三：从数据源关联数据表

AI 探索中心支持将您的腾讯云资源添加为数据源，您可通过自然语言分析数据动因、预测数据趋势。

操作步骤

1. 进入 AI 探索中心 数据知识库 ，单击**添加文件**，选择从**数据源关联**；单击**添加数据源**，支持添加新的数据源。
2. 选择数据源类型，单击**下一步**。
3. 配置数据源：单击**确认添加**后，页面可显示该数据源。确认添加后等待约1分钟，将完成 MCP 服务创建和数据源连通性检测。联接数据源到数据知识库后即可在对话页面选择相应数据知识库进行问答。注意数据源创建成功后，才可使用该数据源进行对话问答。

> **说明：**
> 用户名默认为 root。

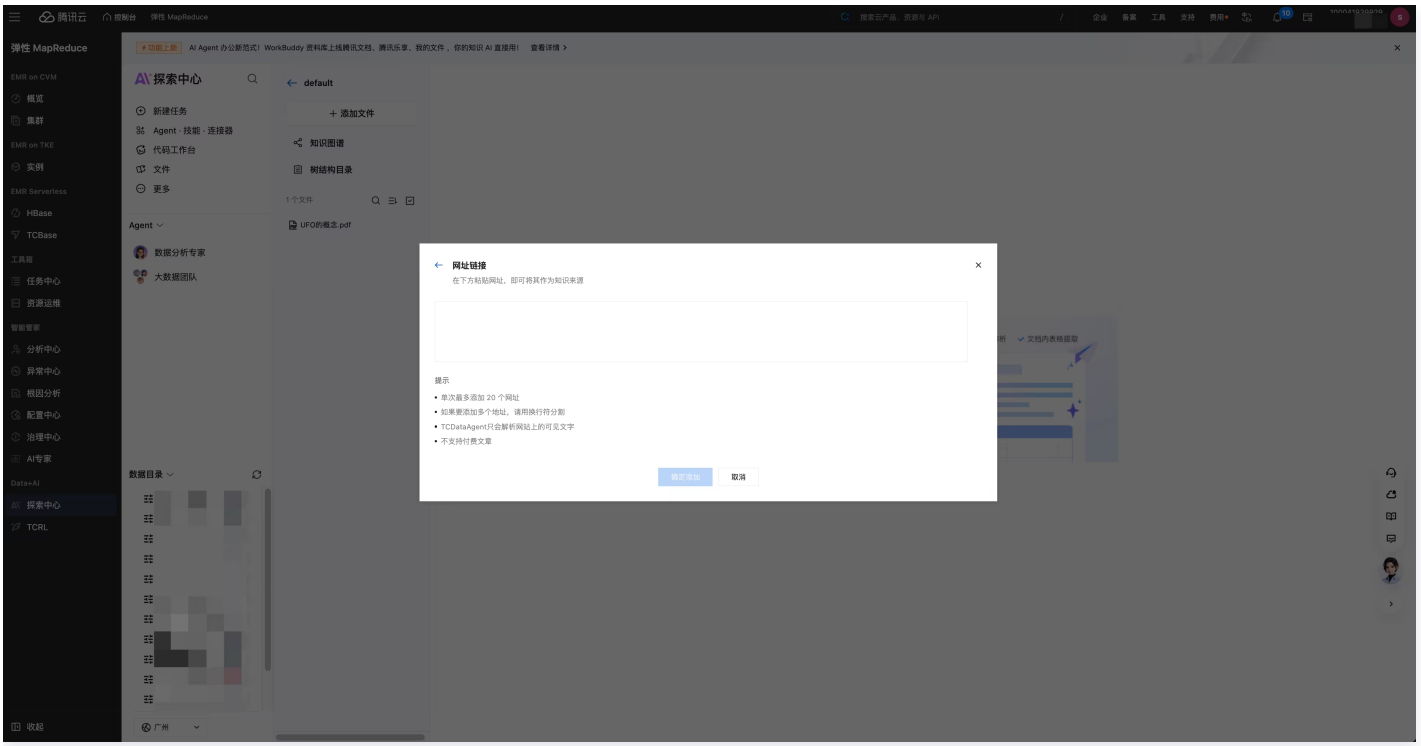


> **说明：**
> 当前支持添加弹性 MapReduce StarRocks、Hive、Trino 数据源、腾讯云数据湖计算 DLC、TCHouse-D、TCHouse-C、腾讯云数据库 MySQL 为数据源，更多其他数据源功能敬请期待。数据湖计算 DLC、数据仓库 TCHouse-D 元数据服务由 多模态智能数据湖 TCLake 提供服务。需要您开通 TCLake 服务，此服务不会产生任何费用。

方式四：从网页链接添加

AI 探索中心支持将网页链接添加为数据源。

> **注意：**
> - 单次最多添加20个网址。
> - 如果要添加多个地址，请用换行符分割。
> - TCDataAgent 只会解析网站上的可见文字。
> - 不支持付费文章。



配置数据源

您可单击数据源操作栏中的**配置**。对数据源的 MCP 服务配置进行修改。支持配置数据源名称、用户名、密码和是否开启数据采样。单击**完成**后，平台将自动更新数据源服务。

查看数据源

进入 AI 探索中心知识库中，单击拟查看详情的数据源卡片。您可在数据源详情页查看该数据源的库、表等数据目录信息。
如需进一步查看数据表详情，可在其操作列单击**查看**。您可在数据表详情中查看数据表描述、数据列描述等详细信息。

移除数据源

进入知识库中，对于拟移除的数据源，单击“···”中的**移除**。

> ⚠ **注意：**
> 移除数据源操作仅在 AI 探索中心中移除，不影响源端数据源中的数据。如确定移除，可单击**确定移除**。

检索配置

当前选择知识库后，默认检索配置为“混合搜索”。目前 AI 探索中心支持“向量搜索”、“全文搜索”、“混合搜索”三种模式。用户可点击“数据知识库”后，在“检索配置”中按需调整。

- **向量搜索：**支持“召回数量”（即知识库向大模型召回的最大段落数据，1~20）调整，默认开启“结果重排”（即根据用户的问题意图对召回的文档分段列表进行重新排序）。
- **全文搜索：**支持“召回数量”（即知识库向大模型召回的最大段落数据，1~20）调整，默认开启“结果重排”（即根据用户的问题意图对召回的文档分段列表进行重新排序）。
- **混合搜索：**支持调整“向量搜索”与“全文搜索”权重配置（0~1），支持“召回数量”调整，默认开启“结果重排”。

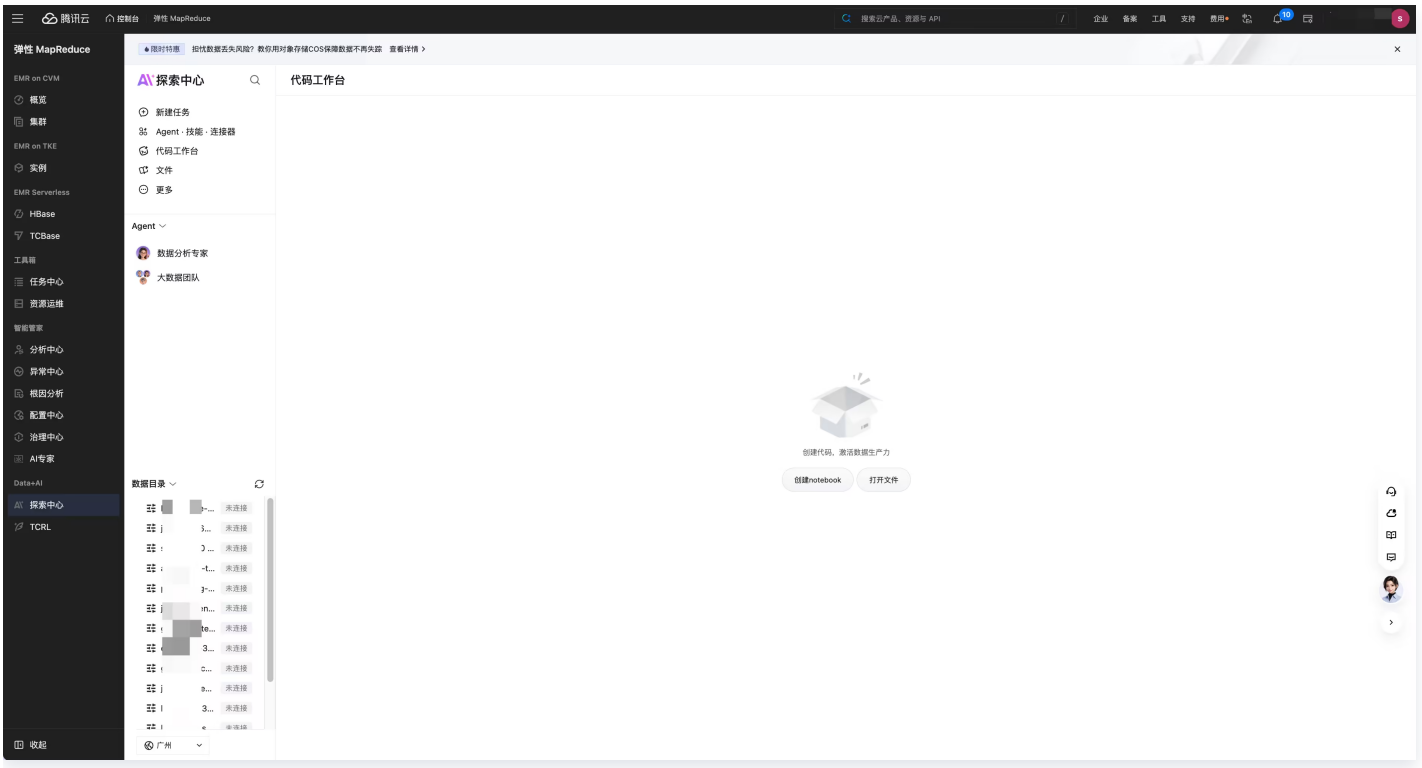
版权所有：腾讯云计算（北京）有限责任公司

代码工作台

最近更新时间：2026-11-20 11:22:03

使用场景

代码工作台 是弹性 MapReduce（EMR）AI 探索中心下的交互式开发与分析环境，支持通过创建 Notebook 或打开文件快速开展代码编写、调试与分析工作，并与 Agent、数据目录及 AI 探索中心功能联动，创建代码，激发数据生产力。



使用指南

使用前配置 Kernel

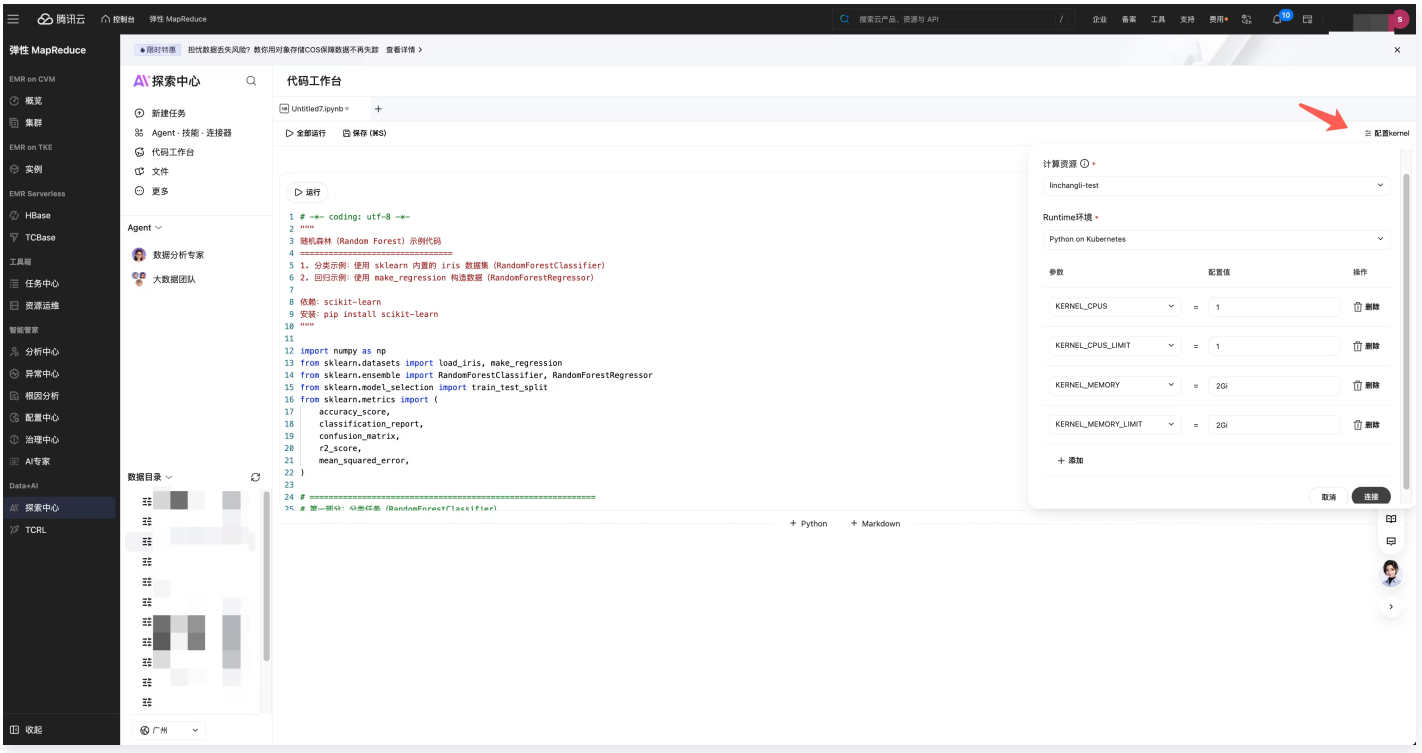
在运行 Notebook 代码前，需先完成 Kernel 配置，为当前交互式会话分配计算资源、指定运行环境并设置资源参数。

1. 打开 Kernel 配置面板：在 Notebook 编辑页右上角点击 配置 Kernel 按钮。
2. 选择计算资源与 Runtime 环境：
   - 计算资源：选择用于执行当前 Kernel 的目标计算资源。
   - Runtime 环境：选择对应的运行时环境（例如 Python on Kubernetes）。

> **说明：**

> 选择计算资源时：请在连接 EMR 集群时请手动打开内网 CLB 并提前安装 EG。

3. 配置 Kernel 参数：在参数列表中按需调整资源请求与上限（支持点击 + 添加 补充参数或点击右侧图标删除多余项）：
   - KERNEL_CPUS：Kernel 的 CPU 资源请求值（如 1）。
   - KERNEL_CPUS_LIMIT：Kernel 的 CPU 资源上限值（如 1）。
   - KERNEL_MEMORY：Kernel 的内存资源请求值（如 2Gi）。
   - KERNEL_MEMORY_LIMIT：Kernel 的内存资源上限值（如 2Gi）。
4. 连接并验证生效：确认参数无误后点击 连接（若无需修改可点击 取消）。连接成功后，Kernel 环境显示正常运行状态。

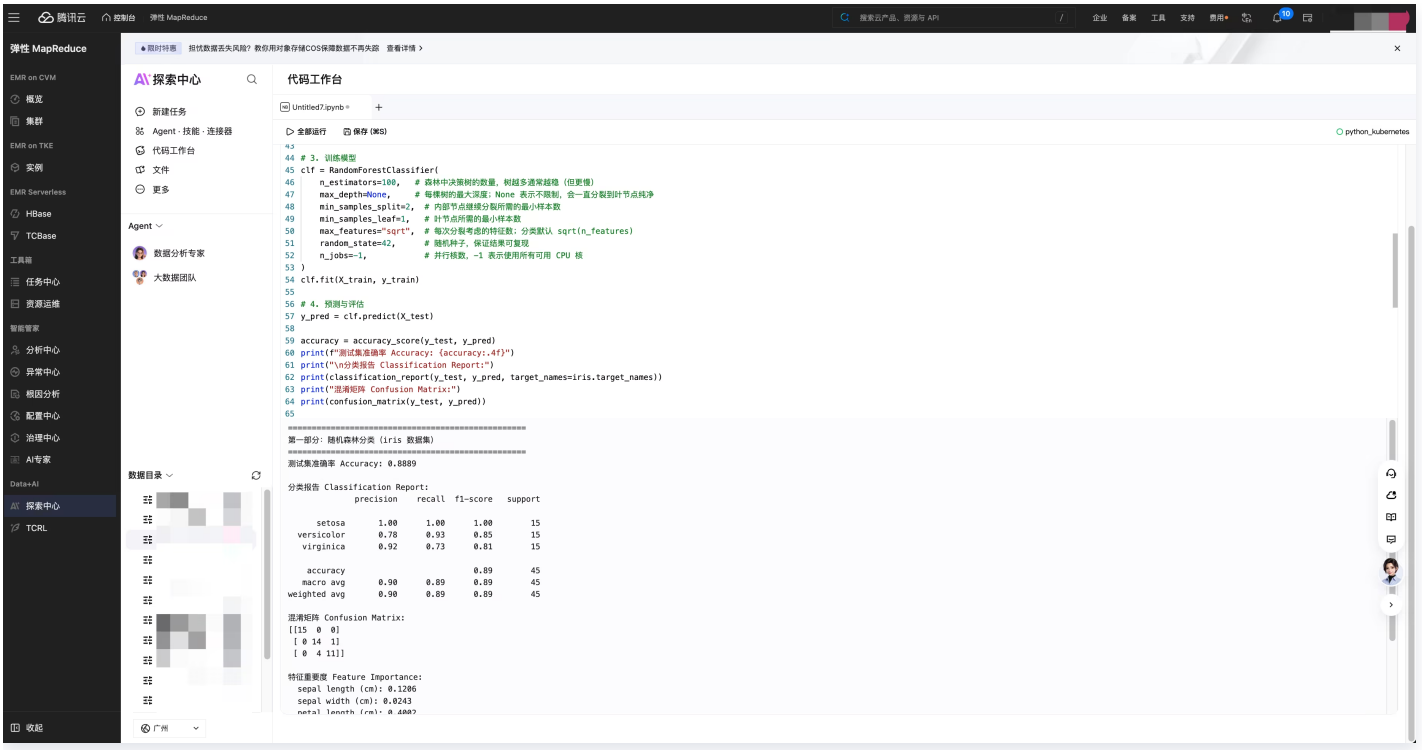


代码开发与调试

在 Kernel 连接成功（右上角显示当前运行时状态，如 python_kubernetes）后，即可在 Notebook 编辑区编写代码、执行运行与调试并查看输出结果。

- 编写与调试代码：在 Notebook 单元格中编写脚本与注释，按需分步执行或点击工具栏的 全部运行 执行整个 Notebook，并通过 保存 保留编辑进度。
- 模型训练与评估示例：以随机森林分类（RandomForestClassifier）为例，可在代码中配置超参数（如 n_estimators、max_depth、min_samples_split、min_samples_leaf、max_features、random_state 与 n_jobs），调用 fit 完成模型训练，再通过 predict 进行推理与评估。

- 查看运行与调试输出：执行完成后，直接在单元格下方查看输出结果与评估指标，辅助排查逻辑与调优参数，例如：
  - 整体精度指标：如测试集准确率（Accuracy）。
  - 分类明细报告：如各类别的精确率（precision）、召回率（recall）、f1-score 及样本数（support）。
  - 混淆矩阵与特征重要性：如混淆矩阵（Confusion Matrix）及各特征的重要性得分（如 sepal length (cm)、sepal width (cm) 等）。



API 调用

最近更新时间：2026-11-20 11:22:03

AI 探索中心底层应用腾讯云数据分析智能体（Tencent Cloud Data Agent）能力，调用 API 能力详情见腾讯云数据分析智能体 API 文档的 API 概览 及相关文档。

EMR 连接器与 Skill 最佳实践

最近更新时间：2026-11-20 11:22:03

AI 探索中心在腾讯云数据分析智能体（Tencent Cloud Data Agent）的数据源与 Agent 技能体系之上，提供面向 WorkBuddy 与外部 Agent 的**连接器**与**技能（Skill）**。用户可通过自然语言分析 EMR 集群的资产、作业、调度、资源与运维状态，使用户无需记忆 API 名称、无需登录多台控制台，即可用一句话完成集群盘点、慢查询定位、资源水位分析与巡检报告生成。

操作场景

面向腾讯云弹性 MapReduce 用户，在 WorkBuddy 中提供一个自然语言驱动的 EMR 运维助手，帮助用户完成六大场景的运维操作：

- **集群发现与资产盘点**
- **健康巡检与异常定位**
- **作业与队列分析**
- **资源调度诊断与调整**
- **自动扩缩容策略治理**
- **成本与变更辅助**

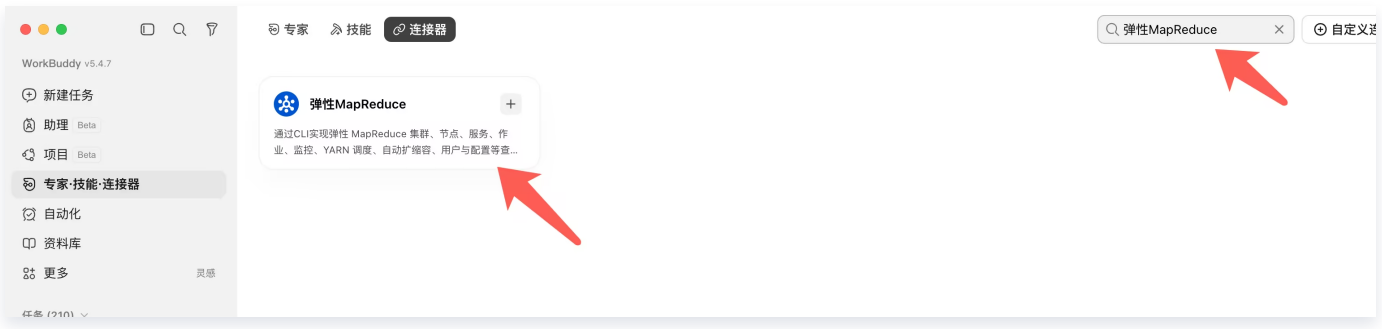
用户不需要进入控制台逐个页面排查，而是可以直接在 WorkBuddy 中提问：

> “帮我检查生产 EMR 集群最近 24 小时是否有资源瓶颈，如果 YARN 队列排队严重，给出优化建议，并在我确认后调整自动扩缩容策略。”

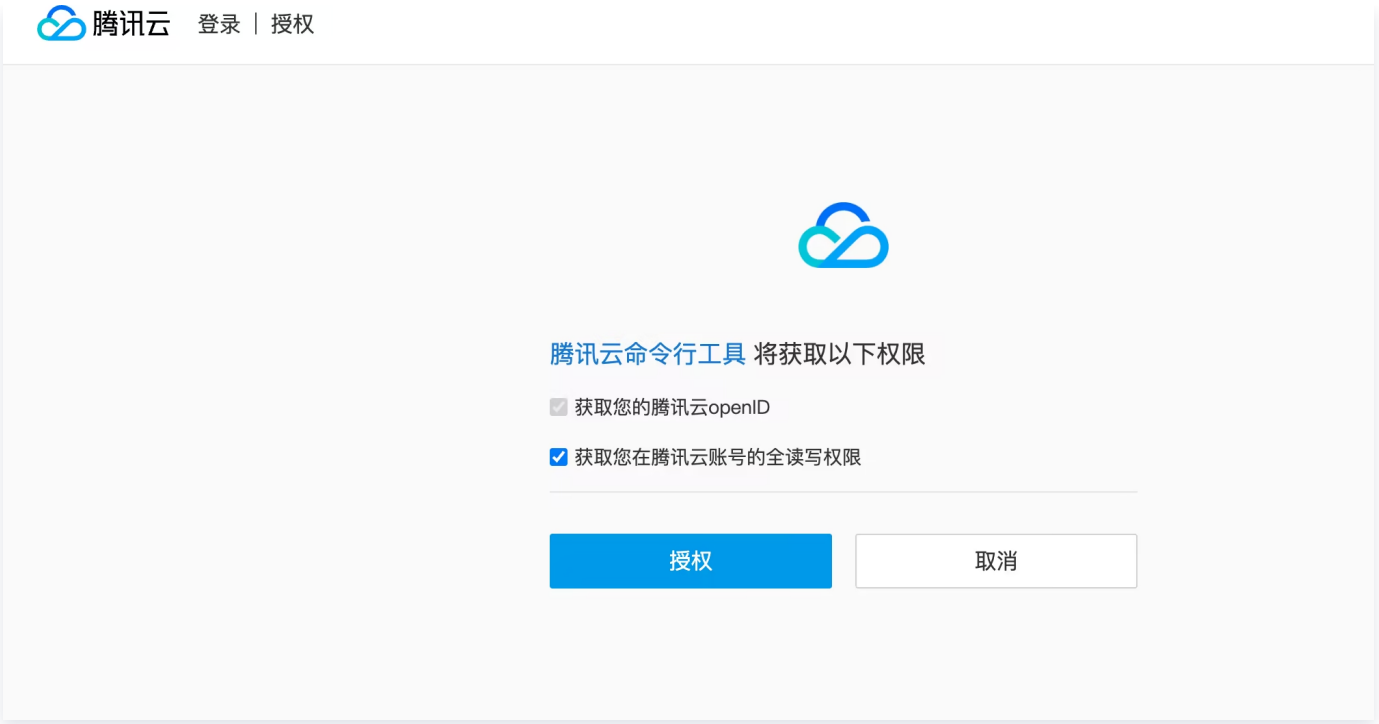
操作步骤

使用方式一

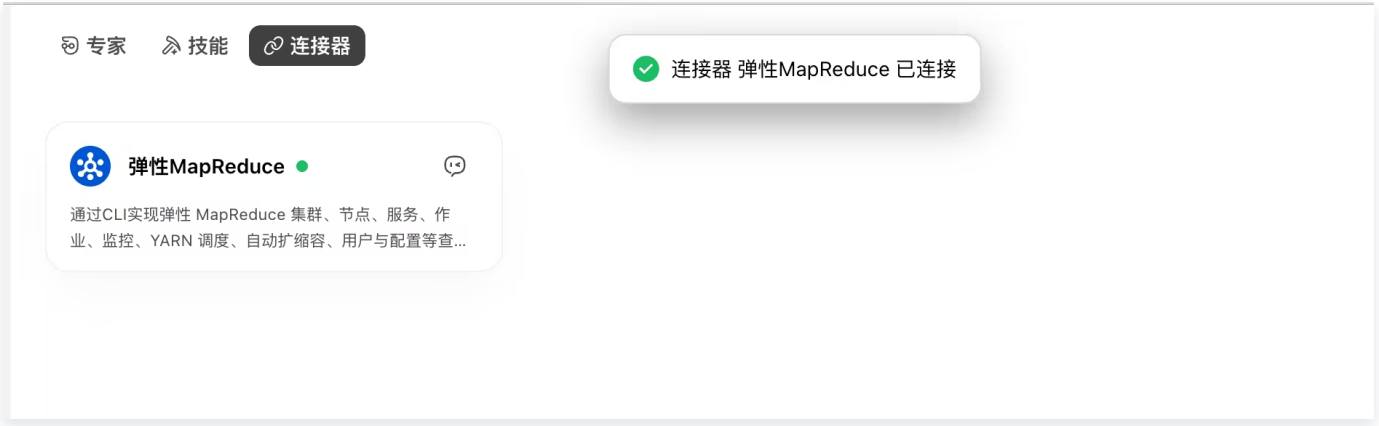
1. 在 WorkBuddy 中点击【专家 · 技能 · 连接器】，选择【连接器】，在搜索框中 输入“弹性 MapReduce“，点击”+“。



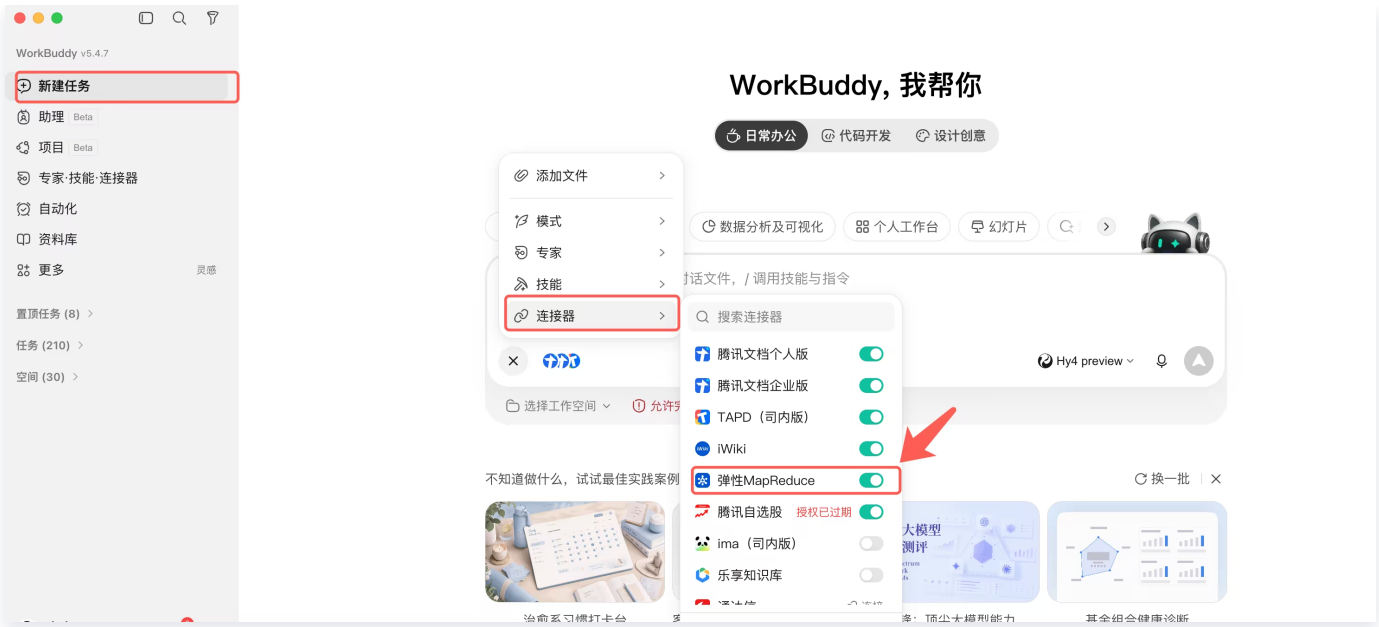
2. 安装授权后：



安装成功：



3. 在对话中，选择【连接器】，勾选“弹性MapReducee“即可。



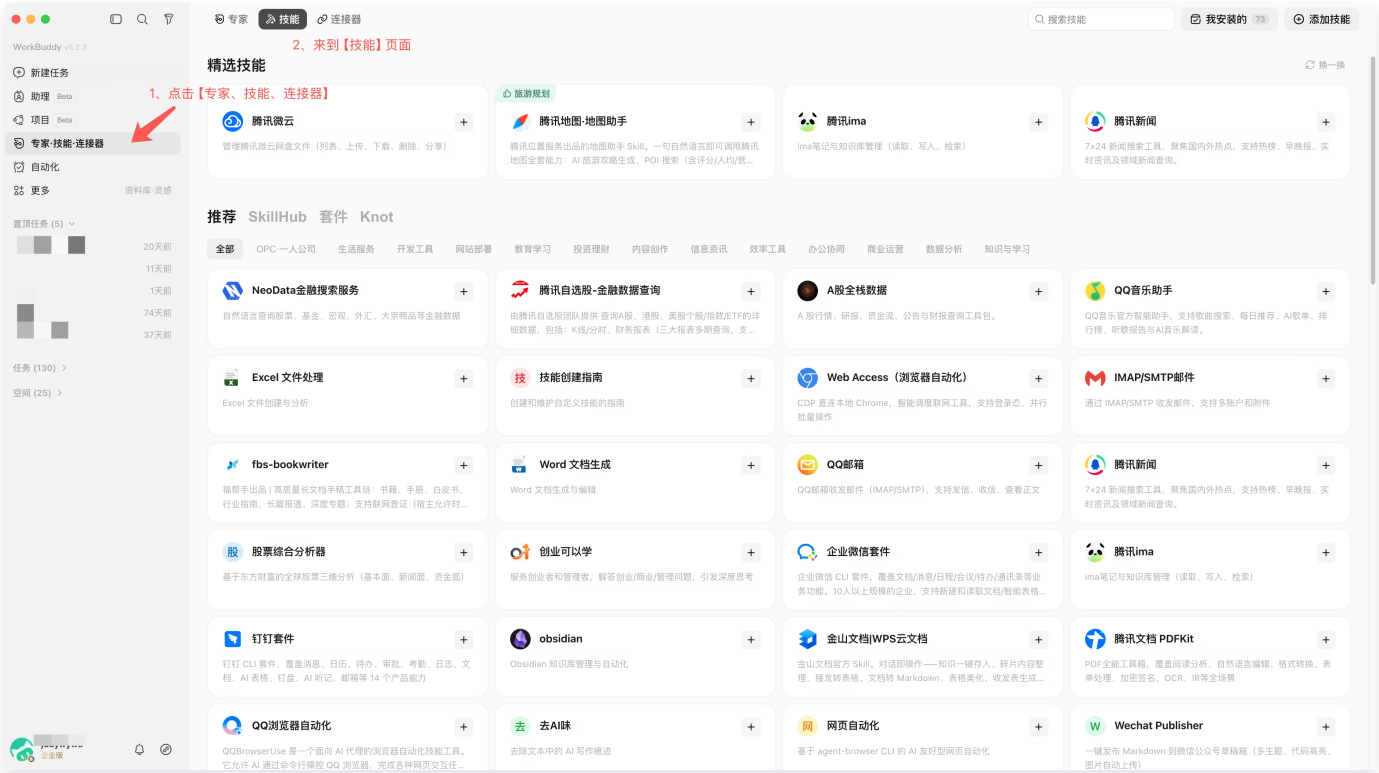
使用方式二

可直接安装的 Skill，可接入 WorkBuddy 或用户自定义 Agent。可通过 提交工单 或联系您的售前获取。

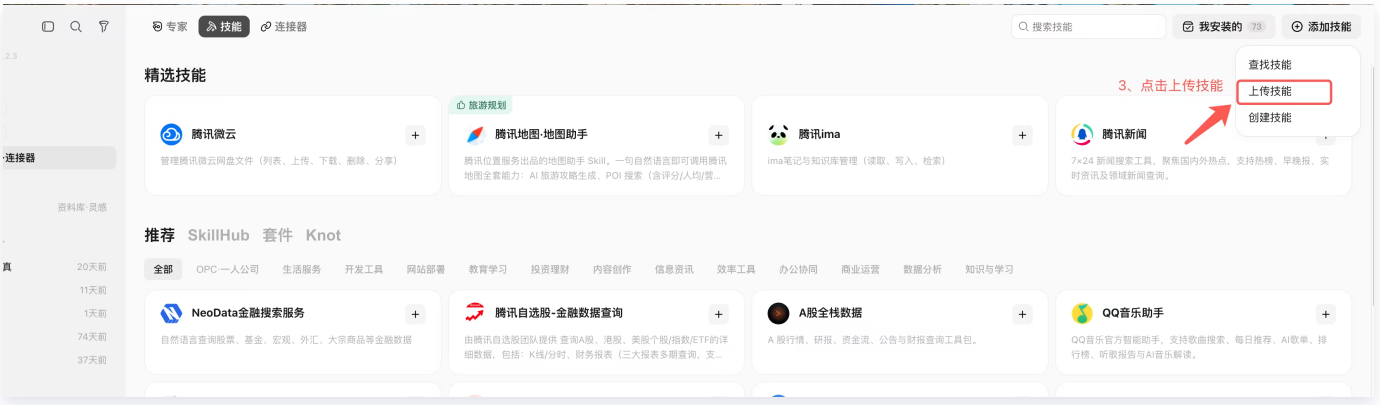
附件： `emr-query.zip`（EMR 查询 Skill 安装包，可接入 WorkBuddy 或用户自定义 Agent，在 WorkBuddy 中上传后即可使用）。

在 WorkBuddy 上传 emr-query-skill 的步骤如下：

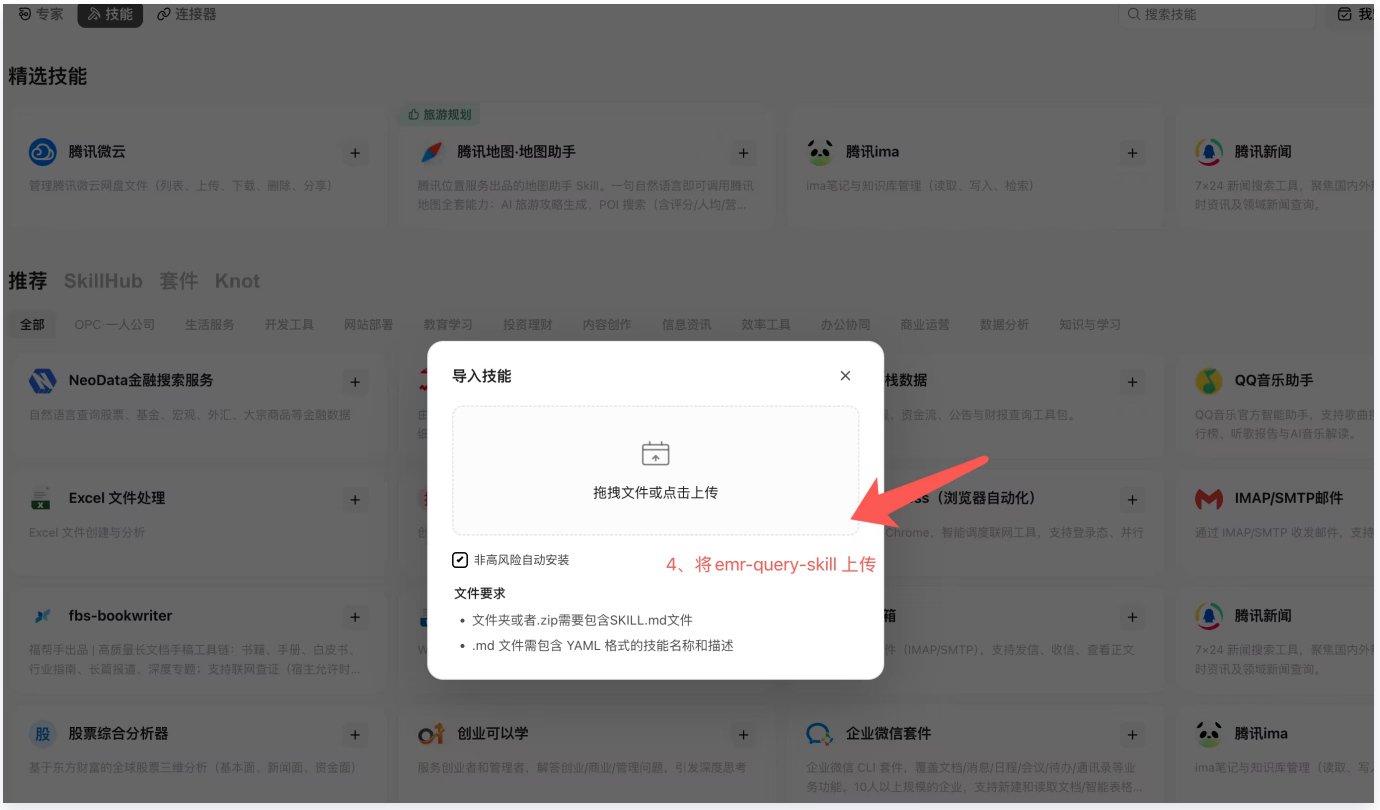
1. 在 WorkBuddy 中点击【专家 · 技能 · 连接器】。
2. 选择【技能】。



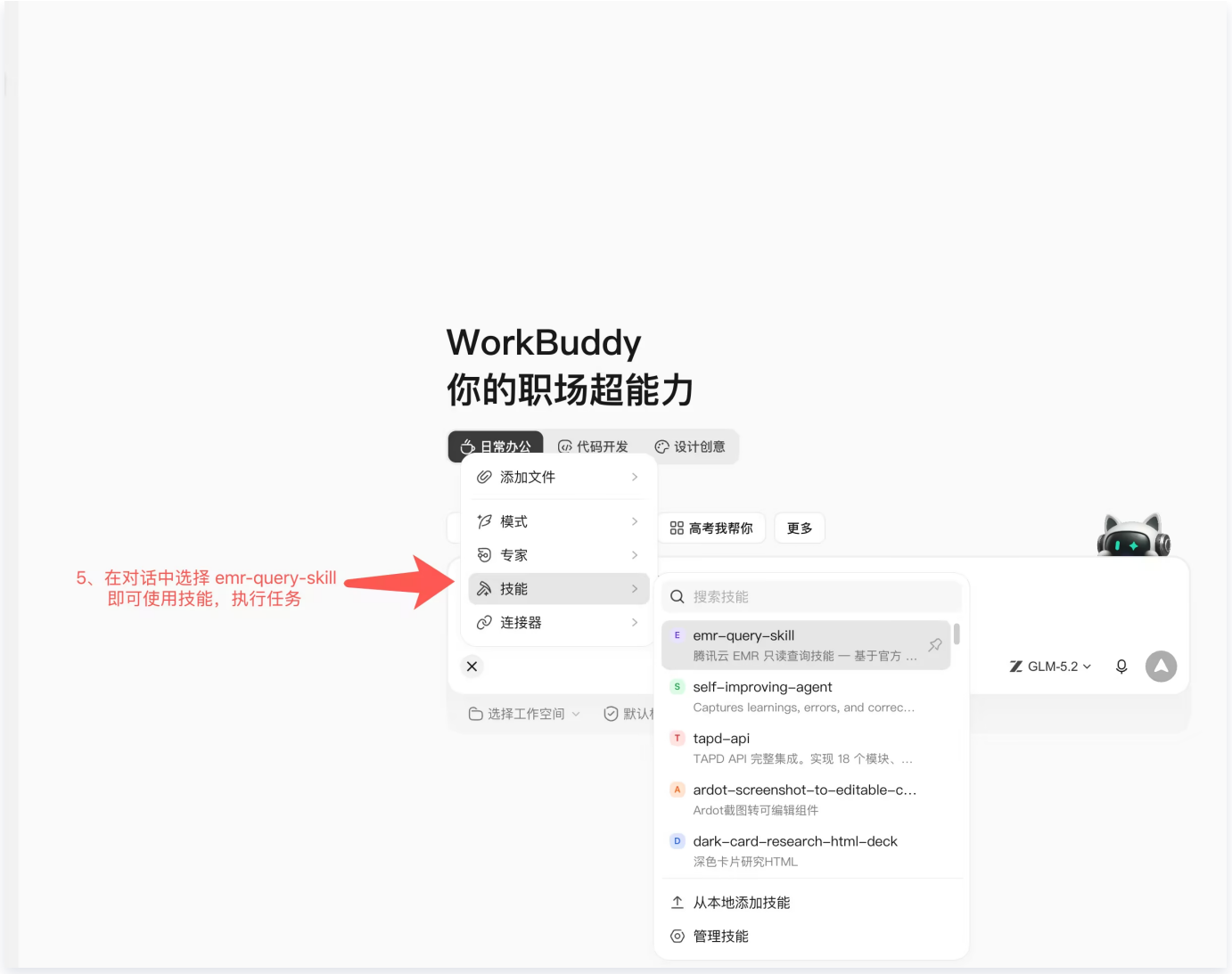
3. 点击**上传技能**。



4. 通过拖拽文件或点击上传 emr-query-skill。



5. 在对话中选择 emr-query-skil，即可使用技能，执行任务。



在使用前请先安装 TCCLI，参考以下文档，或可以通过自然语言让 WorkBuddy 帮您安装，详情可参考 TCCLI 安装文档 。

功能简介

1. 集群查询与资产盘点

用户输入：

> “列一下我广州地域下所有 EMR 集群，重点看生产集群的节点规格、节点数量和服务状态。”

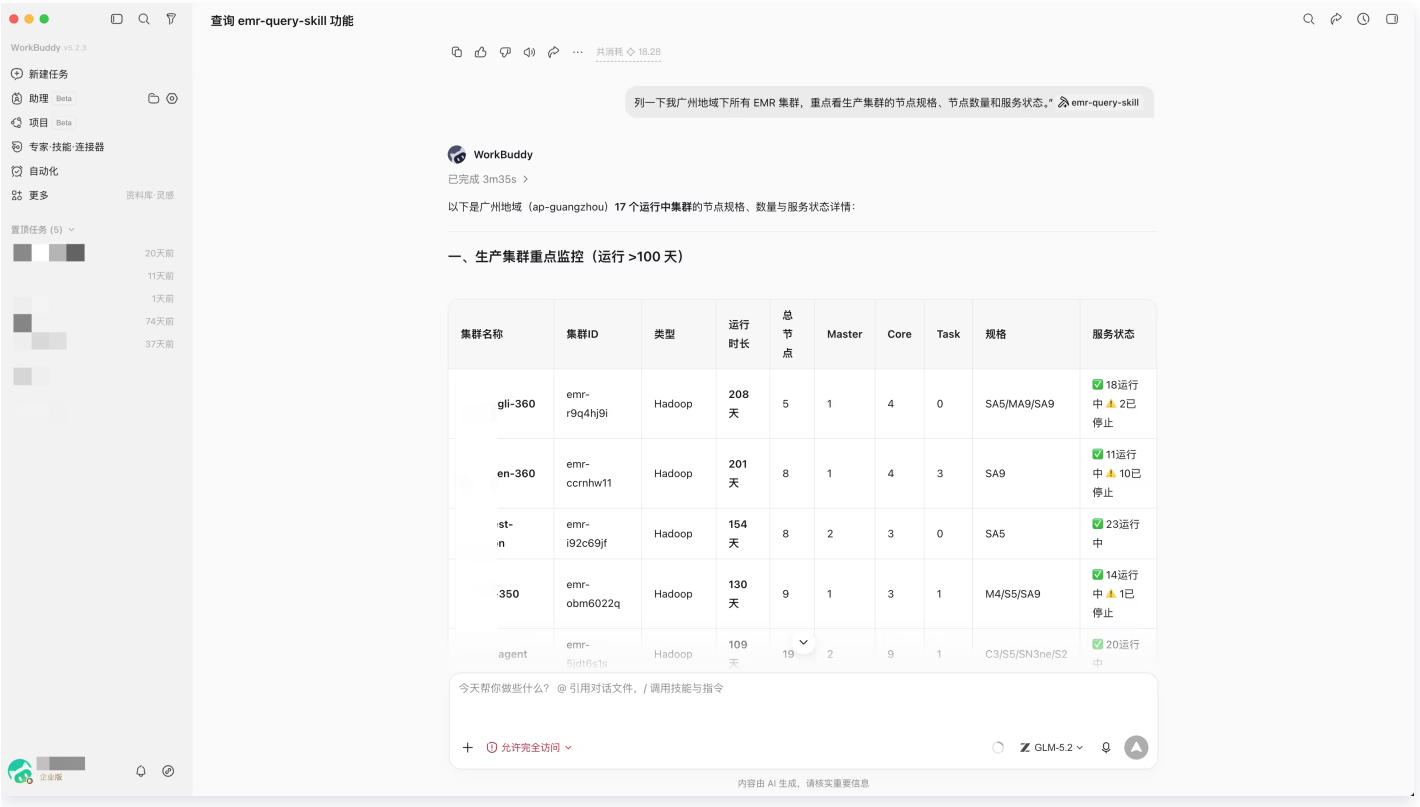
WorkBuddy 调用接口：

- DescribeInstancesList
- DescribeInstances
- DescribeClusterNodes
- DescribeServiceNodeInfos
- DescribeNodeSpec
- DescribeNodeResourceConfigFast

- DescribeCvmQuota

输出内容：

- 集群列表
- 集群状态
- Master / Core / Task 节点数量
- 节点规格
- 服务进程状态
- CVM 配额是否足够



2. 健康巡检与异常定位

用户输入：

> “这个集群最近是不是有异常？帮我看一下监控、事件、巡检结果和操作日志。”

WorkBuddy 调用接口：

- DescribeEmrOverviewMetrics
- DescribeEMREventList
- DescribeInspectionTaskResult
- DescribeInstanceOplog
- DescribeClusterFlowStatusDetail
- DescribeHDFSStorageInfo

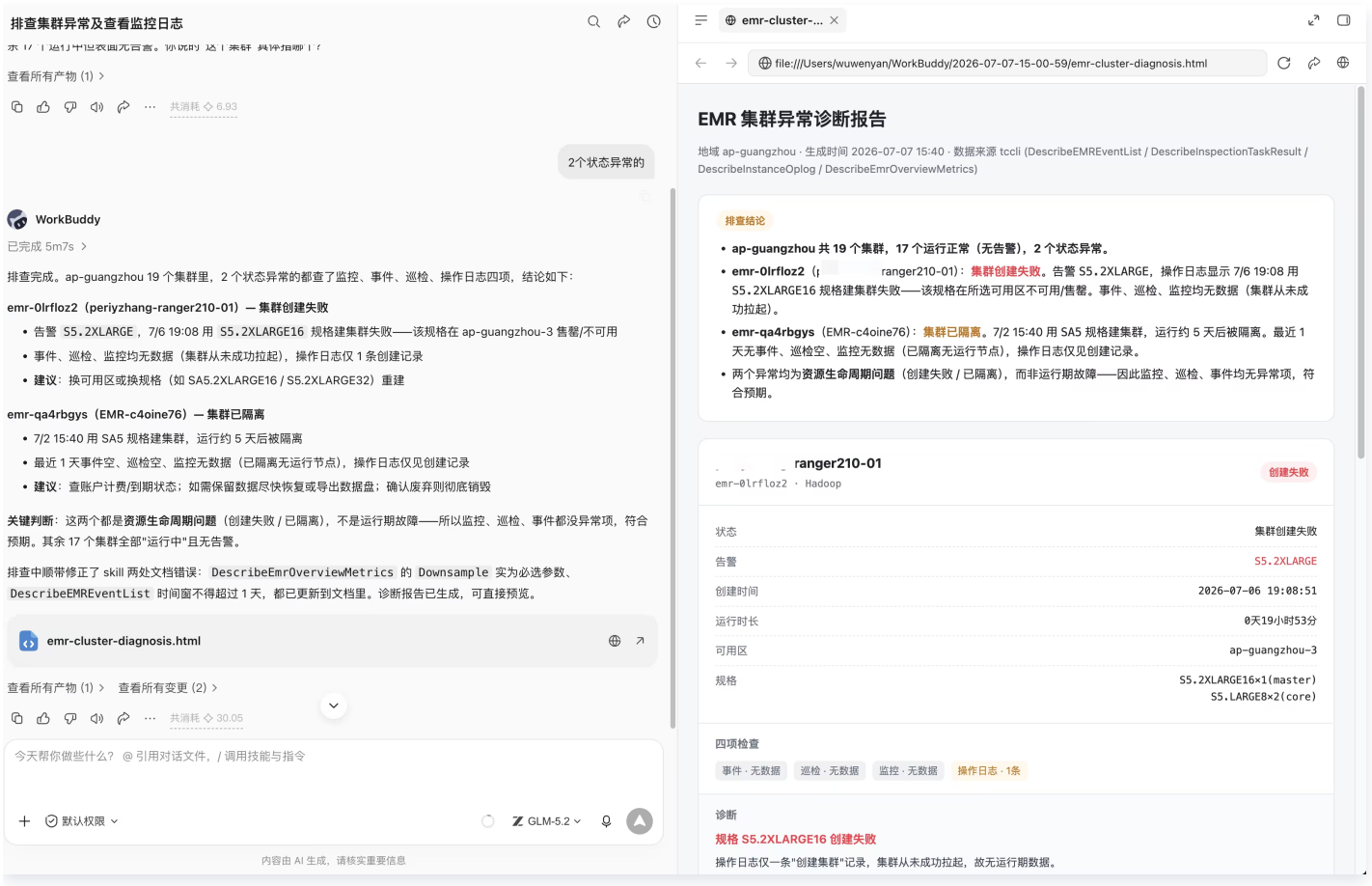
- DescribeInsightList

输出内容：

- CPU / 内存 / 存储使用趋势
- HDFS 文件数量和容量情况
- EMR 事件
- 巡检异常
- 操作日志
- 自动洞察结果
- 风险摘要与优先级







集群异常报告（点击即可查看）：【WorkBuddy】emr-cluster-diagnosis.html 【WorkBuddy】emr-cluster-diagnosis.html

3. 作业与队列分析

用户输入：

> “有哪些集群有在执行Spark任务？是否有慢 SQL？帮我分析 Spark、Hive、Trino 的任务情况。”

WorkBuddy 调用接口：

- DescribeYarnApplications
- DescribeSparkApplications
- DescribeSparkQueries
- DescribeHiveQueries
- DescribeImpalaQueries
- DescribeTrinoQueryInfo
- DescribeKyuubiQueryInfo
- DescribeStarRocksQueryInfo
- DescribeDAGInfo
- DescribeEmrApplicationStatics

输出内容：

- 慢任务列表
- 失败任务列表
- 长时间排队任务
- 高资源消耗用户 / 队列
- Spark / Hive / Trino 查询明细
- DAG 关键路径
- 可能原因：资源不足、队列配置不合理、数据倾斜、并发过高等

版权所有：腾讯云计算（北京）有限责任公司

查询 emr-query-skill 功能

有哪些集群sql跑的慢？为什么今天下午任务跑得慢？帮◆◆◆分析 Yarn、Spark、Hive、Trino 的任务情况。

WorkBuddy

已完成 6m11s >

以下是对今天下午（7月7日 00:00-15:02）广州地域所有 EMR 集群 Yarn、Spark、Hive、Trino 任务情况的完整分析：

一、今天任务执行总览

引擎	有任务的集群	任务总数	说明
YARN	1个（jixinliu-agent）	1个	eval_oom_test0707，9秒，**FAILED**
Spark SQL	0个	0条	所有14个集群均无查询
Hive	1个（periyzhang-ranger230-01）	16条	5条ERROR + 11条FINISHED，均为简单DDL
Trino	0个	0条	所有集群均无查询

二、今天任务详情

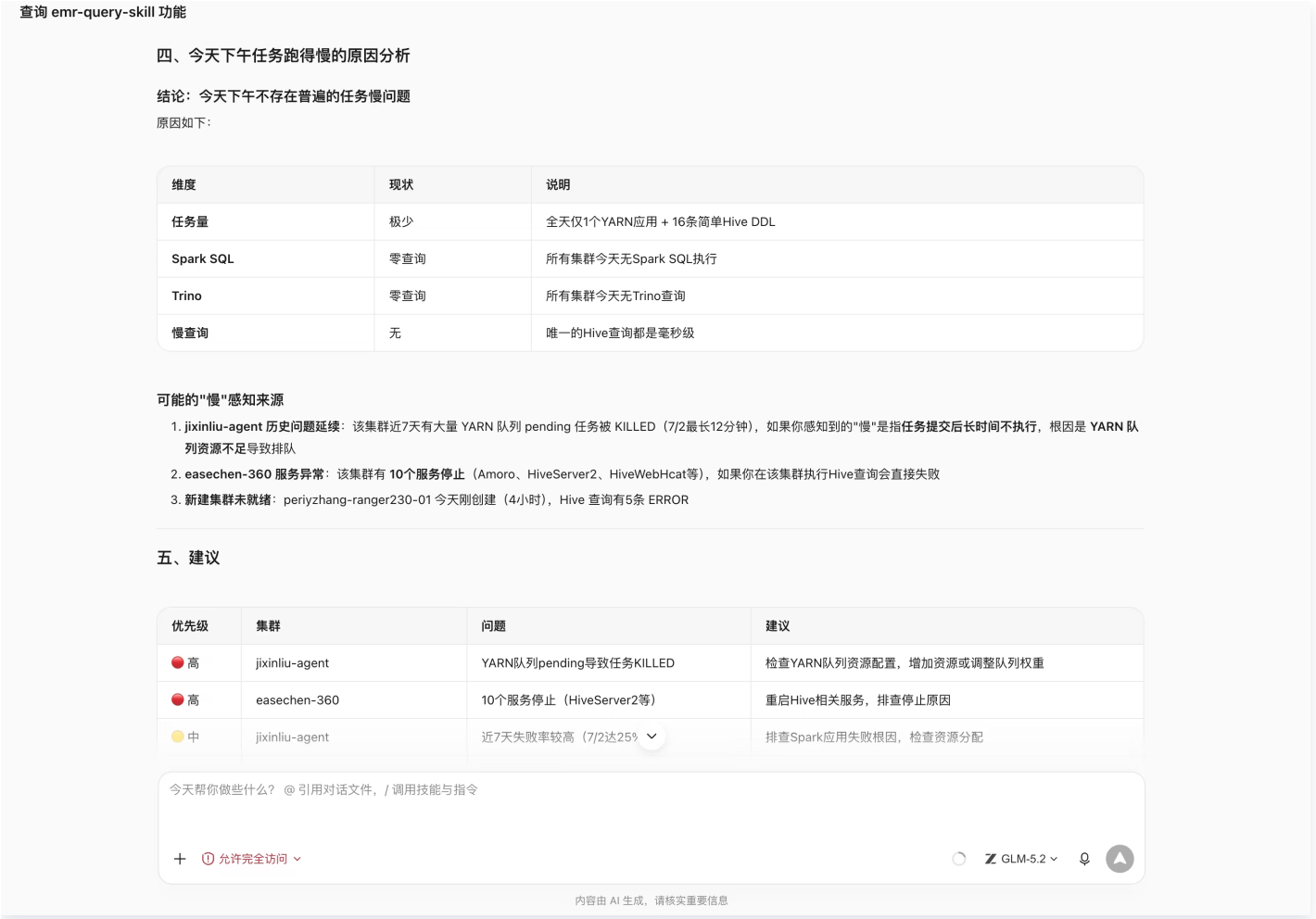
YARN 应用

集群	应用名	执行时间	状态	类型	问题
jixinliu-agent	eval_oom_test0707	9秒	FINISHED/**FAILED**	MAPREDUCE	⚠ OOM测试失败

Hive 查询（periyzhang-ranger230-01）

今天帮你做些什么？ @ 引用对话文件，/ 调用技能与指令

+ 允许完全访问 GLM-5.2



4. 资源调度诊断与调整

用户输入：

> “xx 集群 yarn 对清配置如何？帮我看看是不是资源配置太低。”

WorkBuddy 调用接口：

- DescribeGlobalConfig
- DescribeYarnQueue
- DescribeResourceScheduleDiffDetail
- DescribeYarnScheduleHistory
- DescribeServiceConfGroupInfos

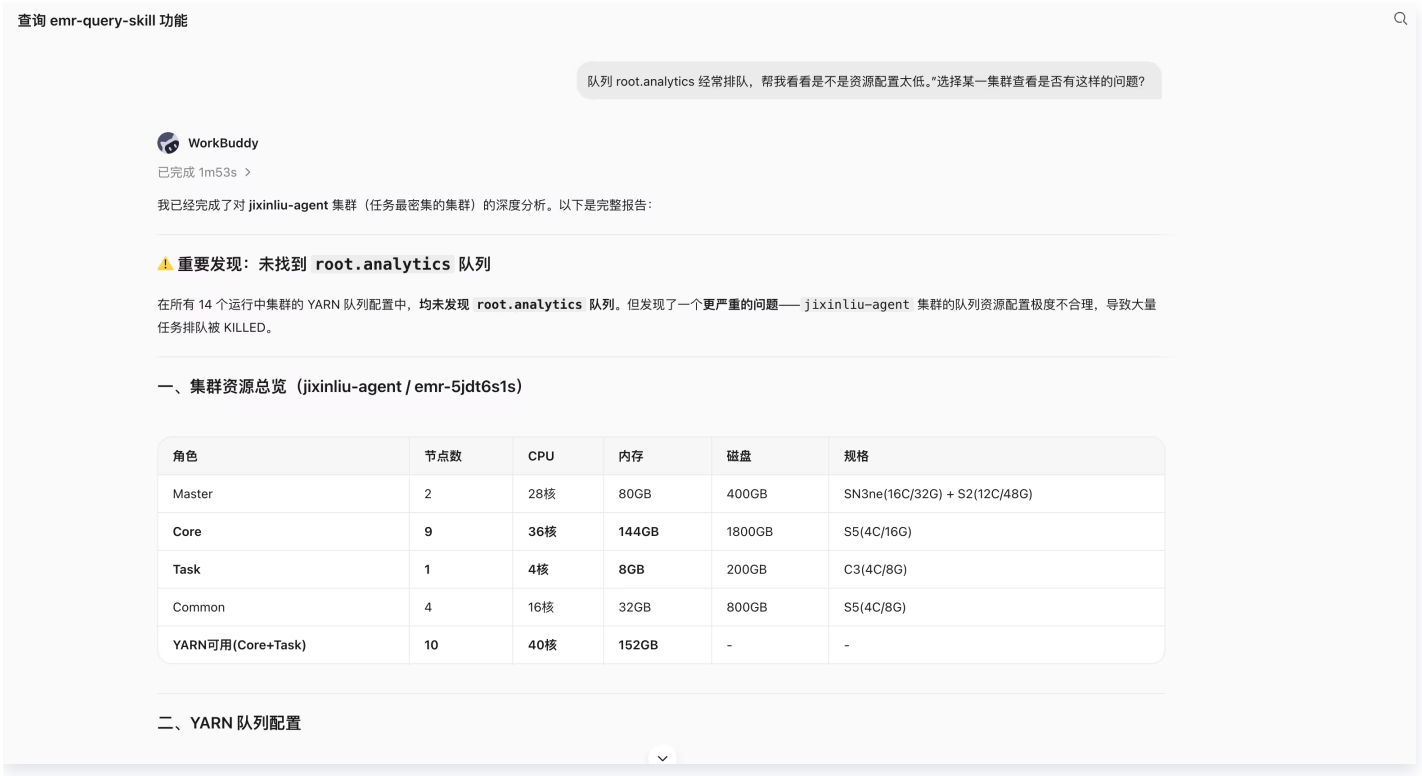
如果需要调整，在用户确认后调用：

- ModifyYarnQueueV2
- DeployYarnConf
- ResetYarnConfig

输出内容：

- 当前调度器配置

- 队列容量
- 最大资源限制
- 用户 / 队列资源占用情况
- 慢任务分析
- 历史变更记录
- 建议调整方案



二、YARN 队列配置

队列	分配内存	分配CPU	占YARN可用比例	状态
root.default	140,087 MB (136.8GB)	40核	90% 内存 / 100% CPU	运行中
root.eval_test	128 MB (0.1GB)	1核	0.08% 内存 / 2.5% CPU	RUNNING

核心问题

```
root.eval_test: 128MB / 1核  →  极度不足!
  - Spark Application Master 最少需要 1GB 内存
  - 该队列连一个 Spark AM 都无法启动
  - 所有提交到该队列的任务必然失败

root.default: 140GB / 40核  →  几乎独占全部资源
  - 165个应用竞争 40核 CPU
  - 平均每个应用仅 0.24核
  - 导致大量任务 pending → KILLED
```

三、近4天 YARN 任务队列分布

队列	总应用	失败	KILLED	慢任务(>1分钟)	KILLED率
root.default	165	43	89	122	54%
root.eval_test	7	0	7	6	100%
合计	172	43	96	128	-

查询 emr-query-skill 功能

三、近4天 YARN 任务队列分布

队列	总应用	失败	KILLED	慢任务(>1分钟)	KILLED率
root.default	165	43	89	122	54%
root.eval_test	7	0	7	6	100%
合计	172	43	96	128	-

慢任务详情

root.default 队列 TOP5 慢任务（全部被 KILLED）：

任务名	耗时	状态	类型
eval_pending_run-20260705-cd1ca2f8a2_job_1	5.2分钟	KILLED	SPARK
eval_pending_run-20260705-cd1ca2f8a2_job_3	5.2分钟	KILLED	SPARK
eval_pending_run-20260705-cd1ca2f8a2_job_2	5.2分钟	KILLED	SPARK
eval_pending_run-20260705-cd1ca2f8a2_job_4	5.2分钟	KILLED	SPARK
eval_pending_Yarn-02-yarn-queue-pending_job_1	5.1分钟	KILLED	SPARK

root.eval_test 队列（100% KILLED）：

任务名	耗时	状态
eval_gcap_run-20260705-3a977183ba	4.8分钟	KILLED

今天帮你做些什么？ @ 引用对话文件，/ 调用技能与指令

查询 emr-query-skill 功能

任务名	耗时	状态
eval_qcap_run-20260705-3a977183ba	4.8分钟	KILLED
eval_qcap_run-20260705-9a1a494ad8	3.5分钟	KILLED
eval_qcap_run-20260704-550ead2819	3.2分钟	KILLED

四、问题根因分析

为什么任务经常排队?

根因链路：

1. root.eval_test 仅 128MB/1核 → 任务无法启动 → pending → 超时 KILLED
2. root.default 虽有 140GB/40核，但 165 个应用竞争 → 资源碎片化 → pending
3. 任务名都是 eval_pending_* → 明确是队列资源排队测试 → 资源不足被 KILLED
4. KILLED 率高达 54%（default）/ 100%（eval_test）

资源配置问题汇总

问题	严重程度	说明
root.eval_test 资源极度不足	🔴 严重	128MB/1核，Spark AM 最少需 1GB
root.default KILLED 率 54%	🔴 严重	165个应用竞争40核，资源严重不足
无资源预留/弹性配置	🟡 中等	未配置 maximum-capacity 弹性扩容
未设置 user-limit-factor	🟡 中等	单用户可占满队列，导致其他用户饥饿

五、优化建议

查询 emr-query-skill 功能

五、优化建议

短期（立即执行）

优先级	队列	当前配置	建议配置	说明
🔴 P0	root.eval_test	128MB/1核	**32GB/8核**	至少能启动 Spark AM + 1个 Executor
🔴 P0	root.default	140GB/40核	**120GB/32核**	释放部分资源给 eval_test
🟡 P1	root.default	无上限	maximum-capacity=80%	允许弹性借用 eval_test 空闲资源

中期（1周内）

1. **新增 `root.analytics` 队列**（如果你确实需要）：从 root.default 中分出 30% 资源
2. **配置 user-limit-factor**：限制单用户最多使用队列 2 倍资源
3. **启用 maximum-am-resource-percent**：限制 AM 资源占比，避免 AM 占满队列

建议的队列配置方案

```
root (100%)
├── root.default (60%, max 80%)       → 日常任务
│   内存: ~91GB, CPU: 24核
├── root.eval_test (20%, max 40%)     → 测试任务
│   内存: ~30GB, CPU: 8核
└── root.analytics (20%, max 60%)     → 分析任务（新建）
    内存: ~30GB, CPU: 8核
```

> **总结**：`root.analytics` 队列在你当前集群中不存在。真正的问题是 **`root.eval_test` 队列资源严重不足（128MB/1核）** 和 **`root.default` 队列任务过载（KILLED 率 54%）**。建议重新分配队列资源，并考虑新增 analytics 专用队列。需要我帮你查看其他集群的队列配置吗?

查看所有产物 (1) >

5. 自动扩缩容策略治理

用户输入：

> “帮我看一下这个集群有没有设置自动扩缩容，策略是否合理。”

WorkBuddy 调用接口：

- DescribeAutoScaleGroupGlobalConf
- DescribeAutoScaleStrategies
- DescribeAutoScaleRecords
- DescribeNodeDataDisks

如需新增或修改策略，在用户确认后调用：

- AddMetricScaleStrategy
- ModifyAutoScaleStrategy

输出内容：

- 当前伸缩组配置
- 当前自动扩缩容规则
- 最近扩缩容记录
- 是否存在频繁扩缩、策略不生效、阈值不合理等问题
- 推荐策略，例如：
  - YARN 待分配内存持续高于 70% 超过 10 分钟时扩容
  - 空闲 Task 节点持续 30 分钟时缩容
  - 设置最小 / 最大节点数保护

> ⚠ **注意：**
> 截图中标红的 DeleteAutoScaleStrategy 不接入，因此 WorkBuddy 不直接提供删除策略能力，可提供“禁用 / 修改为安全阈值 / 引导控制台处理”等替代方案。


查询easechen-360集群自动扩缩容配置
帮我看一下 n-360这个集群有没设置自动扩缩容，策略是否合理。 emr-query-skill
WorkBuddy
已完成 3m36s
查完了，i-360（emr-cc 1），广州六区，带 Ray 的 Hadoop 集群）的自动扩缩容情况如上图，核心结论一句话：
框架搭了，规则没配 — 弹性能力空转
伸缩组是有的（group-aqmc3o04，边界 1–20 节点、优雅缩容已开启、按量付费），但负载策略和定时策略全是空的，历史扩缩容记录 0 条，当前弹性节点 0 台。等于只建了个壳，从来没触发过。
策略是否合理?
项
状态
评价
触发规则
负载 0 条 + 定时 0 条
❌ 核心问题，伸缩组形同虚设
上限 20 台
20×SA9 = 80核/320GB
⚠️ 偏大，峰值成本易失控，建议收紧到 5–10
优雅缩容
已开启
✅ 合理，缩容前迁移容器避免中断
按量付费
弹性节点 POSTPAY
✅ 合理，与固定节点包年包月互补
固定节点到期
8 台全部 7/29 到期
ℹ️ 续费窗口内可用弹性补短期算力
建议
1. 补触发规则：最关键。典型做法是负载策略——YARN 可用内存 < 30% 扩容、> 70% 缩容，冷却 5–10 分钟
2. 收紧上限：从 20 降到 5–10，按实际峰值定
今天帮你做些什么？ @ 引用对话文件，/ 调用技能与指令
默认权限
GLM-5.2


查询easechen-360集群自动扩缩容配置
建议
1. 补触发规则：最关键。典型做法是负载策略——YARN 可用内存 < 30% 扩容、> 70% 缩容，冷却 5–10 分钟
2. 收紧上限：从 20 降到 5–10，按实际峰值定
3. 固定节点 7/29 就到期，如果想保留弹性缓冲，现在正好把规则配上
需要我帮你看下 YARN 队列配置或监控水位，定一个具体的扩缩容阈值吗?
展示详情
-360 集群自动扩缩容分析：伸缩组框架已建但无任何触发策略，自动扩缩容实际未生效
结论：伸缩组已建框架，但未配置任何触发策略 — 自动扩缩容实际未生效
集群存在一个弹性伸缩组（边界 1–20 节点、优雅缩容已开启），但负载策略与定时策略均为空，历史扩缩容记录为 0，当前弹性节点数为 0。弹性能力处于「空转」状态。
伸缩组配置（已创建）
组名 group-aqmc3o04
节点下限 1
节点上限 20
优雅缩容 已开启
付费方式 按量付费
当前弹性节点 0 台
触发策略（缺失）
负载策略 0 条
定时策略 0 条
历史扩容记录 0 条
历史缩容记录 0 条
无规则 → 伸缩组无法被触发，形同虚设
集群节点概览（固定节点）
Master
1 台
SA9 8核/32GB
Core
4 台
SA9 4核/16GB
Task
3 台
SA9 4核/16GB
付费 / 到期
包年包月
2026-07-29
固定算力合计 8 台 = 28 vCPU / 96GB（带 Ray、Spark、Kyuubi、Impala、HBase 等服务）
策略合理性评估与建议
今天帮你做些什么？ @ 引用对话文件，/ 调用技能与指令
默认权限
GLM-5.2



6. 成本与变更辅助

用户输入：

> “目前EMRr 相关 sa5机型价格？哪些机型性价比最高？ “

WorkBuddy 调用接口：

- InquiryPriceUpdateInstance
- InquiryPriceScaleOutInstance
- InquiryPriceRenewInstance
- InquirePriceRenewEmr
- DescribeCvmQuota

如用户确认后，可调用：

- ModifyResource
- ResizeDataDisks
- AttachDisks
- ModifyAutoRenewFlag

输出内容：

- 当前配置
- 目标配置
- 价格差异

- 配额检查
- 影响范围


emr-query-skill 目前emr 相关sa5机型价格？哪些机型性价比最高？
emr-query-skill 目前emr 相关sa5机型价格？哪些机型性价比最高？
WorkBuddy 预计消耗 ◇ 55.25 ~ 531.09
加载技能 emr-query-skill
我来查一下 EMR 相关的 SA5 机型询价。先看看询价接口的具体参数。
定位代码、运行校验：*.md
询价用 InquiryPriceCreateInstance 。先看它的参数，以及查可用机型的 DescribeNodeSpec 。
查看 InquiryPriceCreateInstance.md
DescribeNodeSpec 能返回机型规格（含 SA5）。先确认 tccli 可用与凭证状态，然后拉取广州地域的节点规格，筛选 SA5 机型。
检查 tccli 可用性与凭证
凭证可用。现在拉取节点规格。ZoneId 100007 是广州三区，用较新的产品 ID 查询各机型规格与价格。
查询广州三区节点规格
接口返回了机型规格但没有价格字段。我用 Python 解析出所有 SA5 机型的 CPU/内存规格，然后再对每个 SA5 机型逐一询价。先提取 SA5 机型清单。
确认执行命令?
正在准备执行命令 · 已消耗 ◇ 88.51
今天帮你做些什么？ @ 引用对话文件，/ 调用技能与指令
默认权限
Claude-Opus-4.8

emr-query-skill 目前emr 相关sa5机型价格？哪些机型性价比最高？

已完成 4m13s >

猪猪，SA5 全系我都跑了一遍真实询价（腾讯云 EMR，广州七区，包年包月，含 CVM + EMR 服务费，不含磁盘），结果如下 👇

💰 EMR SA5 机型价格总览（广州七区 · 包月）

机型	规格	内存配比	原价/月	折后/月	折后 元/核/月
SA5.LARGE8	4核8G	1:2	334.88	221.29	**55.32** ⭐
SA5.2XLARGE16	8核16G	1:2	669.76	442.58	**55.32** ⭐
SA5.4XLARGE32	16核32G	1:2	1339.52	885.16	**55.32** ⭐
SA5.8XLARGE64	32核64G	1:2	2679.04	1770.32	**55.32** ⭐
SA5.12XLARGE96	48核96G	1:2	4018.56	2655.47	**55.32** ⭐
SA5.LARGE16	4核16G	1:4	502.32	330.48	82.62
SA5.2XLARGE32	8核32G	1:4	1004.64	660.95	82.62
SA5.4XLARGE64	16核64G	1:4	2009.28	1321.91	82.62
SA5.8XLARGE128	32核128G	1:4	4018.56	2643.81	82.62
SA5.12XLARGE192	48核192G	1:4	6027.84	3965.72	82.62
SA5.16XLARGE256	64核256G	1:4	8037.12	5287.62	82.62
SA5.16XLARGE288	64核288G	1:4.5	8706.88	5724.37	89.44

今天帮你做些什么？ @ 引用对话文件，/ 调用技能与指令

默认权限 Claude-Opus-4.8

emr-query-skill 目前emr 相关sa5机型价格？哪些机型性价比最高？

SA5.16XLARGE256	64核256G	1:4	8037.12	5287.62	82.62
SA5.16XLARGE288	64核288G	1:4.5	8706.88	5724.37	89.44
SA5.32XLARGE576	128核576G	1:4.5	17413.76	11448.74	89.44
SA5.64XLARGE1152	256核1152G	1:4.5	34827.52	22897.48	89.44

🏆 性价比结论

按每核成本看，SA5 呈明显的三档梯度：

- **性价比最高（55.32 元/核/月）** —— 所有 **1核:2G** 配比机型：
  `SA5.LARGE8 / SA5.2XLARGE16 / SA5.4XLARGE32 / SA5.8XLARGE64 / SA5.12XLARGE96`
  这些是**计算密集型/通用大数据负载的首选**，单核最便宜，规格随需求线性放大即可。
- **中档（82.62 元/核/月）** —— 所有 **1核:4G** 内存增强机型，比同核数标配贵约 **50%**。只在跑 Spark/内存计算、需要大内存时才值。
- **最贵（89.44 元/核/月）** —— 超大规格（16X-288 / 32X / 64X），单核成本反而最高，一般只在需要单机超大内存、减少节点数时才用。

给你的选型建议： 常规 EMR 集群优先选 1:2 配比机型（如 Core 节点用 `SA5.8XLARGE64`），单核最省；内存吃紧的场景再上 1:4 机型；除非有强单机诉求，不建议用超大规格机型。

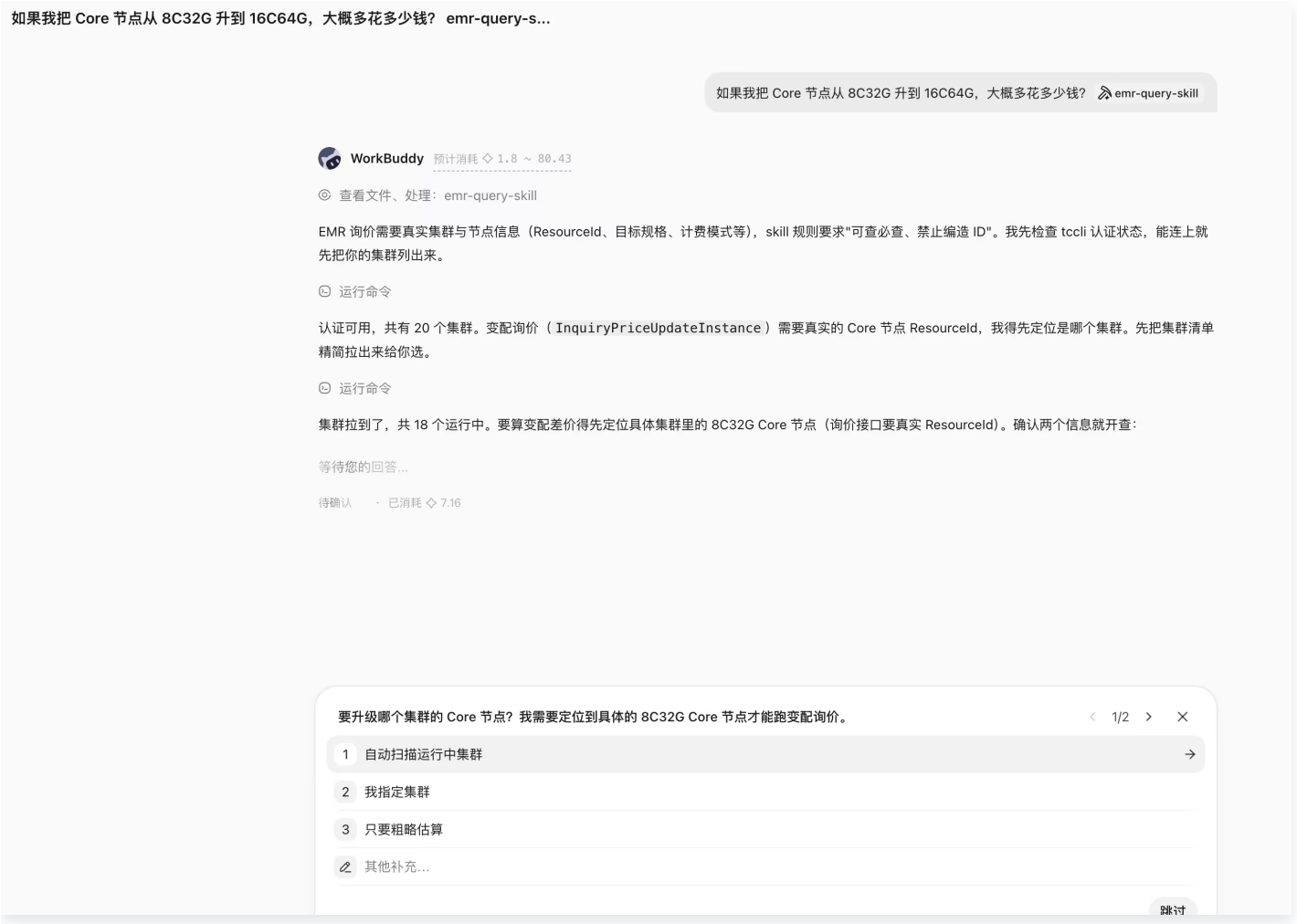
> 📌 说明：以上为节点计算费（CVM + EMR 服务费），不含云盘；折扣按当前包月刊例价（约 6.6 折）。按量计费单价与此成正比、排序一致。如需带磁盘的整机报价或换其他地域/其它机型（SA9、S 系列），跟我说一声即可继续跑。

查看所有变更 (1) >

共消耗 ◇ 195.76 Claude-Opus-4.8 12:02

今天帮你做些什么？ @ 引用对话文件，/ 调用技能与指令

默认权限 Claude-Opus-4.8



接口总览

如您需要调用相关接口，通过您的 Agent 运维管理 EMR 集群，接口可参考，详见 API 文档 。

关键限制：只读（Read-Only）

- **允许：** `Describe*` / `Inquiry*` / `Inquire*` **接口**
- **禁止：** `Create*` / `Modify*` / `Terminate*` / `Run*` / `Delete*` / `ScaleOut*` / `Add*` / `Set*` / `Mount*` / `Uninstall*` / `Install*` / `Sync*` / `Attach*` / `Resize*` / `Deploy*` / `Reset*` / `Convert*` / `StartStop*`

1. 集群资源（10个接口）

典型场景：查看所有集群、查某个集群配置、估算扩容/续费成本。

接口名称	功能说明	复杂度
`DescribeInstancesList`	查询集群列表（含状态、名称、类型、运行时长）	简单

DescribeInstances	查询集群详情（含节点配置、软件信息）	简单
DescribeClusterNodes	查询集群节点列表（按角色过滤）	简单
DescribeCvmQuota	查询 CVM 可用配额	简单
DescribeInstanceRenewNodes	查询集群续费节点信息	简单
InquiryPriceCreateInstance	创建集群询价	复杂（JSON）
InquiryPriceRenewInstance	续费询价	简单
InquirePriceRenewEmr	EMR 续费询价（新版）	简单
InquiryPriceScaleOutInstance	扩容询价	复杂（JSON）
InquiryPriceUpdateInstance	变配询价	复杂（JSON）

2. 信息查询（23个接口）– 核心作业与监控

典型场景：排查作业失败原因、分析查询性能、查看 HDFS 存储使用、获取巡检建议、追踪事件。

接口名称	功能说明	引擎/场景
DescribeSparkApplications	Spark 应用列表	Spark
DescribeSparkQueries	Spark SQL 查询记录	Spark SQL
DescribeHiveQueries	Hive 查询记录	Hive
DescribeImpalaQueries	Impala 查询记录	Impala
DescribeKyuubiQueryInfo	Kyuubi 查询信息	Kyuubi
DescribeTrinoQueryInfo	Trino 查询信息	Trino
DescribeStarRocksQueryInfo	StarRocks 查询信息	StarRocks
DescribeYarnApplications	YARN 应用列表	YARN
DescribeHDFSStorageInfo	HDFS 存储信息	HDFS

版权所有：腾讯云计算（北京）有限责任公司

DescribeHBaseTableOverview	HBase 表概览	HBase
DescribeHBaseTableRequestMetric	HBase 表请求指标	HBase
DescribeHBaseTableStoreSizeMetric	HBase 表存储大小指标	HBase
DescribeEmrApplicationStatics	EMR 应用统计	全局
DescribeEmrOverviewMetrics	EMR 概览指标	全局监控
DescribeEMREventList	EMR 事件列表	事件/告警
DescribeInstanceOplog	实例操作日志	审计
DescribeInsightList	洞察/建议列表	智能分析
DescribeInspectionTaskResult	巡检任务结果	运维巡检
DescribeClusterFlowStatusDetail	集群流程状态详情	流程/任务
DescribeDAGInfo	DAG 执行信息	工作流
DescribeNodeResourceConfigFast	节点资源配置快速查询	节点配置
DescribeNodeSpec	节点规格查询	规格查询
DescribeServiceNodeInfos	服务节点信息	服务分布

3. 扩缩容（4个接口）

典型场景：查看自动扩缩容历史、分析节点磁盘使用。

接口名称	功能说明
DescribeAutoScaleStrategies	查询自动扩缩容策略
DescribeAutoScaleRecords	查询自动扩缩容执行记录
DescribeAutoScaleGroupGlobalConf	查询扩缩容组全局配置
DescribeNodeDataDisks	查询节点数据盘信息

4. 集群服务（2个接口）

典型场景：查看调度器配置、全局参数。

接口名称	功能说明
DescribeGlobalConfig	查询全局配置
DescribeResourceSchedule	查询资源调度配置（如 Capacity/Scheduler）

5. YARN 调度（2个接口）

典型场景：查看队列资源分配、对比调度配置变更。

接口名称	功能说明
DescribeYarnQueue	查询 YARN 队列配置
DescribeResourceScheduleDiffDetail	查询调度配置差异详情

6. 用户管理（2个接口）

典型场景：查看集群用户权限、用户组配置。

接口名称	功能说明
DescribeUsersForUserManager	查询用户列表（用户管理）
DescribeGroupsSTD	查询用户组信息

7. 配置管理（1个接口）

典型场景：查看各服务的配置参数（如 Hive、Spark、HDFS 配置）。

接口名称	功能说明
DescribeServiceConfGroupInfos	查询服务配置组信息

8. Serverless HBase（2个接口）

典型场景：查看 Serverless HBase 实例状态与配置。

接口名称	功能说明
DescribeSLInstance	查询 Serverless HBase 实例详情
DescribeSLInstanceList	查询 Serverless HBase 实例列表

9. 其他（2个接口）

接口名称	功能说明
`DescribeJobFlow`	查询 JobFlow（作业流）
`DescribeYarnScheduleHistory`	查询 YARN 调度历史

调用方式

方式一：简单参数（直接展开）

适用于标量参数、无嵌套对象：

```
tccli emr <Action> --region <region> --version 2019-01-03 --cli-unfold-
argument --Param1 value1 --Param2 value2
```

示例：查询集群列表

```
tccli emr DescribeInstancesList --region ap-guangzhou --version 2019-01-
03 --cli-unfold-argument --DisplayStrategy clusterList --Limit 100
```

方式二：复杂参数（JSON 文件）

适用于 `Filters`、`FlowParam`、`ResourceSpec` 等嵌套结构：

```
# 第1步：生成参数骨架
tccli emr <Action> --generate-cli-skeleton > /tmp/<Action>.json

# 第2步：编辑 JSON 填入真实参数

# 第3步：调用
tccli emr <Action> --region <region> --version 2019-01-03 --cli-input-
json file:///tmp/<Action>.json
```

示例：查询集群流程状态

```
tccli emr DescribeClusterFlowStatusDetail --generate-cli-skeleton >
/tmp/DescribeClusterFlowStatusDetail.json
# 编辑 JSON 后执行：
```

```
tccli emr DescribeClusterFlowStatusDetail --region ap-guangzhou --
version 2019-01-03 --cli-input-json
file:///tmp/DescribeClusterFlowStatusDetail.json
```

认证方式

方式	命令	说明
OAuth 登录	`tccli auth login`	浏览器扫码登录，自动持久化凭证
已有凭证	`~/.tccli/default.credential`	复用已登录的凭证

健康检查（验证认证是否有效）：

```
tccli emr DescribeInstancesList --region ap-guangzhou --version 2019-01-
03 --cli-unfold-argument --DisplayStrategy clusterList --Limit 1
```

错误指南

错误码前缀	含义	建议处理
`AuthFailure.*`	CAM 权限不足	授权 `emr:<Action>` 权限
`UnauthorizedOperation`	资源级未授权	检查 `qcs::emr` 资源权限
`InvalidParameter.*`	参数格式/范围错误	执行 `help --detail` 修正
`MissingParameter`	缺少必传参数	用 skeleton 或接口文档补齐
`ResourceNotFound.*`	资源不存在	检查集群 ID/节点 ID 是否有效
`FailedOperation`	集群类型不匹配	检查集群类型、调度器前提
`LimitExceeded`	频率限制	减少并发，稍后重试

每个接口文件包含：功能描述、必选参数、可选参数、tccli 示例、实测状态、错误码。

快速上手示例

查询集群列表

```
tccli emr DescribeInstancesList --region ap-guangzhou --version 2019-01-
03 --cli-unfold-argument --DisplayStrategy clusterList --Limit 100
```

查询集群节点

```
tccli emr DescribeClusterNodes --region ap-guangzhou --version 2019-01-
03 --cli-unfold-argument --InstanceId emr-xxxx --NodeFlag all
```

查询 Spark 作业

```
tccli emr DescribeSparkQueries --region ap-guangzhou --version 2019-01-
03 --cli-unfold-argument --InstanceId emr-xxxx --StartTime 1751277600 --
EndTime 1751364000 --Offset 0 --Limit 20
```

查询巡检结果

```
tccli emr DescribeInspectionTaskResult --region ap-guangzhou --version
2019-01-03 --cli-unfold-argument --InstanceId emr-xxxx
```

查询 YARN 队列

```
tccli emr DescribeYarnQueue --region ap-guangzhou --version 2019-01-03 -
-cli-unfold-argument --InstanceId emr-xxxx --Scheduler capacity
```

版权所有：腾讯云计算（北京）有限责任公司