

目录

目录	1
产品简介	2
名词解释	2
行业场景与资源	2
数据联邦查询分析	2
丰富的 API 生态，互相借力，协同发展	2
轻松实现系统集成，规范化、标准化	2
产品优势	2
统一数据控制	2
API全生命周期管理	2
灵活的数据缓存机制	2
统计监控及操作审计	2
产品功能	2
交互查询	2
API开发	3
统计监控	3
明细查询	3
API开发	3
服务管理	3
服务发布	3
应用及授权管理	3
策略管理	3
服务配置	3
调用统计及明细	4
调用统计	4
明细查询	4
API分页说明	4
API调用说明	4
自定义函数	5
资源管理	5
函数管理	5
常见问题	5

产品简介

金山云智·数据服务产品服务于各大应用系统，支持以简单易用的方式和低学习成本的标准SQL，查询多源异构数据，可通过可视化方式完成数据API的开发，并通过API网关实现对API的全生命周期管理，保障API服务的高效注册、发布、开放，安全的将大数据平台的数据能力以 API 的形式开放给合作伙伴，以快速盘活企业数据，全面实现数据资产化、服务化。另外，API网关提供请求加密、身份认证、权限管理、流量控制等多重手段保证 API 安全，提供 API 发布、授权、下线等全生命周期管理，并提供便捷的监控、分析等能力，实时监控API的调用情况。

名词解释

- 交互查询

交互查询服务，支持开发/运维人员方便的对底层数据进行查看和分析，支持单库单实例查询、或跨库联邦查询

- 单实例查询

即每次查询只查询单个数据库中单表的数据，单实例查询支持：Oracle、MySQL、GP、ElasticSearch、HBase、Hive六种数据源

- 跨库查询

即支持关联查询不同数据库中的数据，跨库查询支持：Oracle、MySQL、GP、ElasticSearch、HBase、Hive六种数据源之间的跨库查询

- 服务开发

服务开发支持将底层数据能力封装成标准API接口并发布出来，每个API可以发布多个版本

- API配置

即配置API的路径、请求方式、返回类型、API描述、入参、出参等信息，API保存并配置后才可以注册发布至API网关

- 服务发布

服务发布即将API服务注册至API网关中，发布至API网关的API可以被各应用调用，每个API可以发布多个版本

- API网关

API网关为数据服务提供服务治理的能力，能够为数据类服务提供开放给大数据平台外部的能力。

- 应用

用户需要创建 应用 作为调用 API 时的身份

- AppKey

每个应用都有这样一对密钥对，加密计算后放入请求中作为签名信息

- 授权

授予某个 APP 调用某个 API 的权限，由 API 服务方完成。APP 被授权后才能调用 API。

- IP访问策略

用于应用，设置可以访问/不可以访问应用的IP或IP段

- 流量控制策略

用于 API 服务方对 API 、应用内的APP 按天、小时、分钟、秒进行流量限制

行业场景与资源

数据联邦查询分析

针对：开发人员、运维人员、数据分析人员、业务人员等，提供“交互查询”模块，对自己有权限的测试/生产环境库表进行查询分析，交互查询支持ANSI 2003标准SQL，可对底层Oracle、MySQL、GREENPLUM、HBase、ES、Hive进行单实例查询，或两两跨库关联查询。

丰富的 API 生态，互相借力，协同发展

用户日益膨胀的、碎片化的需求促使企业不断探索新的商业模式，以解决客户的各类场景化问题。数据服务为开发者提供了可视化的封装标准的 API 的服务，开发者在此将标准化的 API 服务整合进自己的应用，从而衍生出新的应用，新的服务。封装后的API产品，可以发布至API 网关，开放出去，以此促进企业建立商业生态、跨界创新。

- 通过 API 网关将企业核心能力开放给合作伙伴，进行深度合作，协同发展
- 将 API 接入服务市场，以 API 的形式开放数据能力供广大开发者采购使用，产生价值；

轻松实现系统集成，规范化、标准化

- 通过 API 网关对系统间接口进行规范统一，用标准化的接口实现系统集成；
- 快速完成资源整合和管理，消除快速发展造成的冗余和浪费，聚力发展业务。

产品优势

统一数据控制

- 支持精确到列级别的数据权限管理，统一数据权限管控
- 支持对元数据的查询审计
- 支持统计数据的访问热度
- 支持集中式的精确到列级别的数据权限管理

API全生命周期管理

- 提供完整的API管理功能，管理API的全生命周期，并生成API详情，提升 API 管理、迭代效率。支持：API编辑、测试、注册、发布、调用等，支持通过API列表可查看API详情，支持API限流，支持将API授权给各应用。
- 提供精准的流控策略，可以控制API、应用下API的访问频率，流控策略灵活可变，支持对API进行分钟、小时、天的流量控制。

灵活的数据缓存机制

- 提供灵活的缓存机制，可通过统一缓存配置页面，进行数据缓存的配置调整。
- 为提高数据查询效率，支持通过创建缓存表的方式对数据进行的预加工，将加工数据存放在指定数据库引擎中，供API调用时快速查询

统计监控及操作审计

- 提供便捷的监控、报警、分析工具，降低 API 运营、维护成本。记录全平台访问记录，可查看每次调用的具体详情和异常情况，快速定位调用问题。可实现对各API、各应用的调用情况统计，监控各API、应用的调用情况和趋势。
- 保存并展示每个API、应用、流控策略、IP访问策略等内容的审核记录，可方便的查看操作人进行的：新建、编辑、审核、授权、取消授权等动作，进行操作审计。

产品功能

交互查询

支持通过可视化的方式快速预览数据情况，支持对Oracle、MySQL、Greenplum、ElasticSearch、HBase、Hive六种数据源的单实例查询和跨库关联查询，并实时查询反馈结果。

API开发

支持通过可视化的方式进行API开发，通过标准SQL便捷式进行数据查询，通过可视化界面进行API信息配置，支持API的一键测试和一键注册。提供完整的API管理功能，管理API的全生命周期，并生成API详情，提升API管理、迭代效率。支持：API编辑、测试、注册、发布、调用等，支持通过API列表可查看API详情，支持API限流，支持将API授权给各应用。

统计监控

按租户、项目、作业维度，对交互查询、API调用、应用调用情况进行统计，监控各API、应用的调用情况、调用趋势、QPS趋势等。

明细查询

记录调用访问记录，提供可视化的明细查询页面，展示交互查询、API调用的具体明细情况，可便捷了解每次调用的开始结束时间、耗时、审计情况、异常情况、调用链情况等，快速定位调用问题。

API开发

1. 新建主题 主题用于为API进行分类，点击【新建主题】按钮，进行主题创建。
2. 新建API 点击【新建API】按钮，选择归属主题后，可创建API。
3. 选择查询模式 API开发支持“单实例查询”“跨库查询”两种模式，当选择“跨库查询”时，可以选择引用缓存表进行开发。
4. 数据预览 待查询的数据表可通过左侧tab的【数据预览】功能进行提前预览
5. 输入查询SQL及SQL格式化 用户可以在【查询SQL】界面输入查询SQL语句，运行过程中，可实时查看查询状态：已完成、查询中、查询异常 点击【格式化】可对SQL进行格式化操作。
6. 配置 点击【配置】按钮，可配置API的API Path、请求参数、入参、出参等信息。
7. 测试API 配置保存后，点击【测试】按钮，输入参数测试值，即可对API进行测试。
8. 提交生成API 测试无误后，点击【提交】按钮，即可生成一个版本的API。备注：API支持多版本
 - API最新版本未发布注册API网关的：可编辑未发布的API版本；
 - API最新版本以发布注册API网关的，可编辑DEV版本生成新版本

服务管理

1. API管理 服务开发页面提交生成API后，可在【服务管理-API管理】中查看/操作。服务管理支持对API进行【注册API网关】【测试】【下线】【删除】操作：
 - 测试 【测试】用于测试API的可用性。
 - 注册API网关 【预发布】和【已下线】状态的API，可注册API网关。 点击【注册API网关】，将对应版本的API注册至API网关。
 - 下线 【已发布】状态的API进行下线操作。 点击【下线】按钮，下线对应的API版本。
 - 删除 【预发布】和【已下线】状态的API可以进行下线操作。 点击“删除”按钮，删除对应的API版本。
2. 异步服务管理 发布生产的异步服务任务可以在此页面查看状态并进行启停管理。 点击异步服务的名称，可查看服务具体详情。
3. 缓存表管理 交互查询中创建的缓存表，可以在缓存表管理页面进行查询、操作。
 - 点击缓存表名称，可以查看缓存表详情
 - 自动更新失败的缓存表，可点击“手动更新”按钮进行缓存表更新
 - 未被API引用的缓存表，可直接删除；被缓存表引用的API信息可在“关联API列表”中查看，只有将绑定的相关API全部下线后，才可删除缓存表

服务发布

进入API网关的【服务发布】页面，可对API进行版本发布、查看历史操作、查看API详情等操作，发布后，API可授权给各应用，供应用调用。

1. 版本发布

【待发布】【已发布】【已授权】的API均可进行版本发布动作，发布时相关字段说明如下：

- 密文发布：选择“是”则API密文发布，选择“否”则API明文发布（选择加密时，会对API返回内容进行base 64加密，调用API后，再对结果base 64解密即可）
- 待发布版本：即API最新的待发布版本
- 发布类型：常规发布、灰度发布
- 常规发布：即单版本发布，发布的新版本替换之前的旧版本
- 灰度发布：在现有线上版本的基础上，追加发布一个API版本，并通过配置灰度比值，分配两个版本的调用比例
- 线上运行的API版本：即当前线上运行的API版本
- 灰度版本：选择灰度发布时，待追加发布的API版本
- 灰度比值：线上版本和灰度版本的调用比例（如线上版本V1.0，灰度版本V2.0，灰度比值60%，则：60%的调用请求分配给V1.0，40%的调用请求分配给V2.0）

2. API下线

【已发布】状态的API可进行下线操作。

3. 操作审计

点击【历史操作】可查看API的操作历史

应用及授权管理

进入API网关【应用列表】页面，可创建、编辑、删除、查看应用，应用详情中，可查看授权给此应用的API信息。 进入API网关的【授权管理】页面，可将API授权给已有应用。

策略管理

1. 流量控制策略

针对API，可以设置流量访问控制，提供精准的流控策略。可控制API、应用下API的访问频率，流控策略灵活可变，支持进行分钟、小时、天的流量控制。

2. IP访问控制策略

针对应用，可以设置IP访问控制策略，支持对IP或IP网段配置白名单、黑名单，保障应用访问的安全。

服务配置

关于数据查询、API调用过程中，涉及的：测试返回数据量、查询超时时间等参数，可以在“运维中心-数据服务-服务管理”中进行配置。 服务配置页面主要包含两部分内容：① 项目级别的服务调用情况预览 ② 统一查询引擎的参数配置，下面将分别说明。

1. 服务概览 支持查看各项指标的：交互查询、API调用的使用情况。

服务类型	指标	说明
API 服务	API 数量	租户下，所有项目下 API 的总数量
	总调用量	租户下，所有项目下 API 总调用次数（测试+生产）
交互查询	总任务数	租户下，所有项目下交互查询总任务数
	总查询量	租户下，所有项目下交互查询总次数（测试+生产）

2. 服务配置 通过服务配置，可配置数据服务中：缓存、超时、测试返回条数等参数。

参数	通用查询引擎-配置说明	峰值限制
元数据缓存周期 (H)	元数据缓存的周期，24H 以内，建议值：1H	24 小时
查询结果缓存周期 (D)	查询结果缓存的周期，3 天以内，建议值：1 天	3 天
快查询-超时配置 (S)	快查询的超时时间配置，300S 以内，建议值：10s。 快查询即：查询数据量小于 10G 的关系型数据库的单实例/跨库查询	300 秒
慢查询-超时配置 (S)	慢查询的超时时间配置，10800 秒以内，建议值：300s。 慢查询即：非关系型数据库查询，或查询数据量在 10G-100G 之间的关系型数据库查询	10800 秒
快查询-数据量峰值 (G)	快查询最大支持查询 10G 数据。 快查询即：查询数据量小于 10G 的关系型数据库的单实例/跨库查询	10G
慢查询-数据量峰值 (G)	慢查询最大支持查询 100G 数据。	100G

调用统计及明细

调用统计

1. API调用统计

可在API调用统计，查看API调用详情，如调用次数、成功率、失败率、平均耗时等。 点击“API调用趋势”可查看API的调用趋势和QPS波动情况。

2. 交互查询统计

可在交互查询统计页面中，查看交互查询的调用详情，如查询次数、运行环境成功率、失败率、平均耗时等。

明细查询

1. API调用明细

在【API调用明细】中，可查看每次API调用的明细，包括：运行状态、耗时、审计结果等。点击【查看调用链】，可查看调用链情况。

2. 交互查询明细

在【交互查询明细】中，可查看查询的明细，包括：状态、耗时、审计结果、返回条数等。

API分页说明

通过“API开发”页面开发的API默认不分页，但API仅调用关系型数据库时，支持分页，分页支持及调用情况如下文所示。

1. API分页支持情况 API调用的数据源仅涉及关系型数据库Mysql、Oracle、Greenplum时（含单实例查询或跨库查询），API支持分页但默认不分页。 API调用的数据源涉及非关系型数据库HBase、Hive、ES时（含单实例查询或跨库查询），API不支持分页。 查询不同数据源时，API对分页的支持情况如下：

是否支持分页	Oracle	Mysql	Greenplum	HBase	ES	Hive
Oracle	是	是	是	否	否	否
Mysql	是	是	是	否	否	否
Greenplum	是	是	是	否	否	否
HBase	否	否	否	否	否	否
ES	否	否	否	否	否	否
Hive	否	否	否	否	否	否

2. API分页调用参数 当API支持分页时，可以通过配置默认入参“pageSize”和“pageNum”指定返回哪个分页。

入参	API：查询关系型数据库	API：查询非关系型数据库
pageNum	非必填：不输入表示不分页	不支持此参数
pageSize	非必填：不输入表示不分页	不支持此参数
自定义入参	必填，在API开发时，配置的API入参（有默认值时，可不填）	

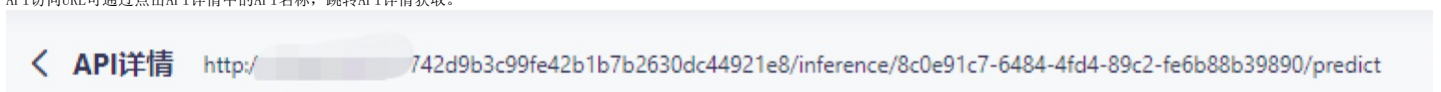
备注1：不分页/不支持分页时，调用返回条数最大10万条，具体可返回条数，可在“运维中心-数据服务-服务配置”中配置

备注2：支持分页时，每页最大返回数为1万条，具体可返回条数，可在“运维中心-数据服务-服务配置”中配置

API调用说明

1. API访问URL

API访问URL可通过点击API详情中的API名称，跳转API详情获取。



2. Header中的传入参数说明

【必选】 ak: 即应用的AppKey, 可在“应用列表-应用详情”中查看。 【必选】 env: 环境: 支持字段online 代表线上环境, test代表测试环境。 【必选】 Content-Type: body参数类型 支持字段 application/json。

3. Body中的传入参数说明

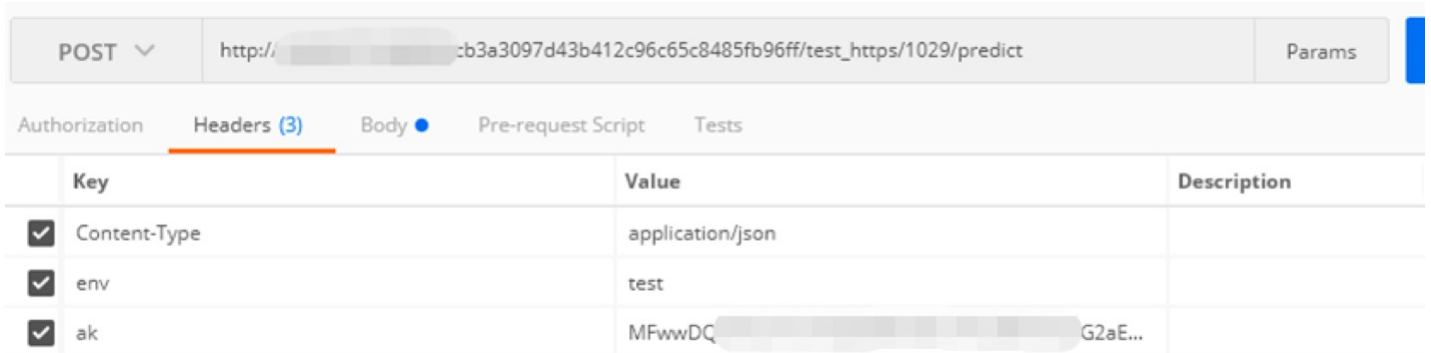
Body采用json格式, 输入参数根据API注册时的入参, 举例如下图所示:

API请求参数:

参数名称	参数类型	参数示例值	参数默认值
aa	Number		

请求body中需填入 { "aa": 5 }

4. POSTMAN 调用示例: 、



自定义函数

资源管理

开发界面上可以进行资源的管理, 数据服务需要使用jar包类的资源, Jar包类只能进行上传更新。

1. 上传资源

- 前置条件: 不存在同名资源, 资源名称符合规范, 资源大小符合规范
- 操作步骤: 开发人员点击创建资源按钮, 填写资源的必要信息, 选择要上传的资源, 点击确定 (备注: 数据服务只能上传JAR类型的资源)。

2. 引用资源

- 前置条件: 当前租户下有可用资源。
- 操作步骤: 在新建自定义函数时, 可引用之前已上传的资源, 具体可参考下文“函数管理”相关说明

3. 资源下载

- 前置条件: 已经存在的资源。
- 操作步骤: 用户进入资源管理页面, 选择某个资源, 点击“变更历史”, 在操作列点击下载。

函数管理

1. 新增函数

用户创建UDF过程中需要保证输入的UDF名称和创建的UDF一致, 来避免覆盖的情况。

- 前置条件: 用户需要使用自定义函数
- 操作步骤: 用户新建函数, 定义函数名, 类型, 用户, 参数说明, 选择依赖的资源等 (备注: UDF租户内共享, 创建后, 各项目都可引用)
- UDF名称: 请输入函数名称, 如: add
- 包名.类名: 依赖的JAR包的: 包名.类名
- 选择资源: 即在“资源管理”中上传的jar包
- 返回值: 请输入返回值的格式, 如: Double
- 用途: 请输入函数用途, 如add(Double a , Double b)表示: 两数相加
- 命令格式: 即函数用法, 如: add(Double a , Double b)
- 参数说明: 如: 银行家舍入法 (0~4: 舍, 6~9: 进, 5的前一位数是偶: 舍, 5的前一位数是奇: 进)

2. 查看函数

操作步骤: 用户点击查看函数, 返回函数的详细信息

常见问题

1. 什么是数据服务?

数据服务产品服务于各大应用系统, 为开发者提供: ① 数据快速查询分析的能力; ② 快速将数据表变成API服务的能力。致力于通过服务封装释放全平台数据能力, 以简单易用的方式和低学习成本的标准SQL统一查询异构存储系统, 盘活企业数据, 全面实现数据资产化、服务化。

2. 数据服务支持查询哪些数据源? Oracle、Greenplum、Hive、ES、HBase、MySQL

3. API是否支持多版本?

API支持多版本发布, 可将不同版本的API发布至API网关, 供各应用调用。

4. API是否支持灰度发布，灰度发布有哪些注意事项？

生产环境支持灰度发布，灰度发布时，注意事项如下：

- API有2个版本以上时可进行灰度发布
- 需保证待灰度发布的不同版本API的入参和出参一致

5. 如何限制API的调用速率？

针对API，可以设置流量访问控制。

- 支持“秒”“分”“时”“天”四种粒度的流控方式
- 支持分别对API总流量和API应用内的流量进行限制
- 可将API批量绑定流控策略（同一个API同时只能绑定一个策略）

5. 是否支持配置黑白名单？

针对应用，可以设置IP访问控制策略：

- 支持白名单（允许）、黑名单（拒绝）两种方式
- 支持将IP、IP段设置为黑/白名单
- 可将应用批量绑定IP访问策略（同一个应用同时只能绑定一个策略）