

LakeFormation

快速入门

文档版本 01
发布日期 2024-01-31



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

目录

1 LakeFormation 使用流程简介.....	1
2 创建 LakeFormation 实例及客户端连接.....	3
3 规划数据湖.....	6
4 MRS 对接 LakeFormation 实例.....	9

1 LakeFormation 使用流程简介

该LakeFormation入门教程介绍了如何创建一个LakeFormation实例并与MRS集群对接，实现统一的数据湖元数据及权限管理。

使用流程简介

MRS与LakeFormation对接的使用流程如下图所示：

图 1-1 LakeFormation 使用流程



约束说明

LakeFormation与MRS集群对接需注意以下约束限制：

- MRS对接LakeFormation前，需要注意以下约束限制：

- MRS集群和LakeFormation实例必须同在一个云账户下且属于同一个Region。
- LakeFormation创建的接入客户端所在虚拟私有云，必须与MRS集群在同一虚拟私有云下。
- MRS集群仅支持对接LakeFormation实例的hive Catalog。
- MRS存量集群需要先完成元数据库和权限策略向LakeFormation实例上迁移，再配置对接。
- **MRS对接LakeFormation后，MRS组件功能约束限制：**
 - Hive暂不支持临时表功能。
 - Hive暂不支持跨集群的列加密表功能。
 - Hive WebHCat暂不支持对接LakeFormation。
 - Hive创建内表时若表目录不为空，则禁止创建表。
 - Hudi表创建前，需要先在LakeFormation上添加Hudi表目录的路径授权，赋予OBS读写权限。
 - Hudi表不支持在LakeFormation管理面编辑表的字段，只能通过Hudi客户端增删改表的字段。
 - Flink读写Hudi场景下同步Hive表，仅支持使用hive_sync.mode=jdbc，不支持hms方式。
 - Spark使用小权限用户登录客户端创建数据库时，若用户没有default库的OBS路径权限，将提示缺少权限，实际创建数据库成功。
- **MRS对接LakeFormation后，权限策略约束限制：**
 - 通过LakeFormation授权仅支持将LakeFormation角色作为授权主体，不支持用户或用户组作为授权主体。
 - PolicySync进程不会修改集群内RangerAdmin Hive模块的默认策略，默认策略仍然生效。
 - PolicySync进程启动后，会与LakeFormation实例的权限进行比对，删除LakeFormation上不存在的非默认策略，请先完成权限策略迁移到LakeFormation实例上。
 - RangerAdmin WebUI界面的Hive模块，禁止执行添加、删除权限非默认策略的操作，统一在LakeFormation实例的数据权限界面进行授权操作。
 - MRS集群取消对接LakeFormation后，RangerAdmin的非默认策略不会清理，需要人工进行清理。
 - Hive暂不支持Grant授权的SQL语句，需统一在LakeFormation实例的数据权限界面进行授权操作。

2 创建 LakeFormation 实例及客户端连接

准备用户

如果您是第一次使用LakeFormation，请参考LakeFormation[准备工作](#)章节，完成用户创建并授权使用LakeFormation等准备工作。

本入门示例，是为了演示使用LakeFormation的全流程，用户还需提前完成以下准备工作：

- 已在IAM服务中创建用于授权的IAM业务用户、用户组，用户已关联至用户组。
- 已在OBS服务中创建一个并行文件系统，并在文件系统中创建用于映射数据表的文件夹。
- 已创建对接LakeFormation所需的相关委托。例如，如果执行MRS对接LakeFormation，则参考[MRS准备工作](#)创建相关委托等。

创建 LakeFormation 实例

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入 LakeFormation控制台。

步骤3 单击页面右上角“立即购买”或“购买实例”，进入购买实例页面。

首次创建实例时界面显示“立即购买”，如果界面已有LakeFormation实例则显示为“购买实例”。

步骤4 按照需求配置以下参数。

表 2-1 创建 LakeFormation 实例参数配置

参数	参数说明	样例
类型	选择实例类型。 <ul style="list-style-type: none">• 共享• 独享	独享


参数	参数说明	样例
区域	选择区域。 不同区域的云服务产品之间内网互不相通。请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。	xxx
计费模式	实例的计费模式。 • 按需收费	按需收费
项目	选择实例所属的项目。	xxx
名称	自定义LakeFormation实例名称。	lakeformation-test
QPS	每秒最大请求数。如果“实例类型”为“共享型”，则无需配置该参数。	10000
企业项目	选择集群所属的企业项目。如果当前无可用企业项目，可以单击“新建企业项目”进行创建。	xxx
描述	当前实例的描述信息。	-
标签	在标签键/值输入框输入内容后单击“添加”，即可添加标签。	-

步骤5 单击“立即购买”，确认配置的相关信息并支付。

步骤6 单击“返回控制台”，在控制台即可查看新创建的LakeFormation实例信息。

说明


创建实例时需要注意配额提醒。当资源配额不足时，建议按照界面提示申请足够的资源，再创建实例。

等待实例状态变为“ 运行中”表示实例已创建成功。

----结束

创建接入客户端

步骤1 登录管理控制台。

步骤2 在左上角单击“”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择**创建LakeFormation实例**已创建的LakeFormation实例（例如lakeformation-test），在左侧导航栏选择“接入管理”。

步骤4 单击“创建客户端”，在弹出的窗口中填写以下参数后，单击“确定”。

如果没有合适的虚拟私有云或子网，可以单击“前往VPC创建”。

表 2-2 创建接入管理客户端参数

参数	说明
客户端名称	自定义接入客户端的名称。
虚拟私有云	待接入客户端所属虚拟私有云。需要与待接入的服务保持一致。
所属子网	待接入客户端所属子网。需要与待接入的服务保持一致。

步骤5 单击“操作”列“查看详情”，可以查看终端节点ID、接入IP、接入域名等信息。

---结束

3 规划数据湖

MRS对接LakeFormation仅支持对接LakeFormation实例的数据目录名称为“hive”的Catalog，名称为“default”的数据库。

LakeFormation实例创建成功后，用户可按照业务规划创建相关Catalog及内部的数据库、表等元数据。

例如参考以下步骤，创建Catalog名称为“hive”（固定名称，不可自定义），并为该Catalog创建数据库“default”（固定名称，不可自定义），数据库内包含两个数据表“table_A”、“table_B”，并进行授权。

创建存储元数据的 OBS 路径

- 步骤1** 创建OBS并行文件系统，详情请参考[创建并行文件系统](#)章节。例如文件系统名称为“lakeformation-test”。
- 步骤2** 在并行文件系统页面，单击已创建的文件系统名称，例如“lakeformation-test”。
- 步骤3** 在左侧导航栏选择“文件”，单击“新建文件夹”，填写待创建的文件夹名称，单击“确定”。继续单击该文件夹名称，单击“新建文件夹”，可以创建其子文件夹。

参考该步骤，依次创建用于存放元数据的路径，例如：

- Catalog存储路径：lakeformation-test/catalog1
- 数据库存储路径：lakeformation-test/catalog1/database1
- 数据表存储路径：lakeformation-test/catalog1/database1/table1、lakeformation-test/catalog1/database1/table
- 函数存储路径：lakeformation-test/catalog1/database1/udf1

---结束

创建 Catalog

- 步骤1** 登录管理控制台。
- 步骤2** 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。
- 步骤3** 在左侧下拉框中选择[创建LakeFormation实例及客户端连接](#)已创建的LakeFormation实例（例如lakeformation-test），在左侧导航栏选择“元数据 > Catalog”。

如果当前实例已包含名称为“hive”的Catalog，则跳过Catalog的创建操作。

步骤4 单击“创建Catalog”，配置以下参数后，单击“提交”。

- Catalog名称: hive
固定名称，不可自定义。
- 选择位置: 单击“+”，选择存储位置，例如选择“obs://lakeformation-test/catalog1”，单击“确定”。
- Catalog类型: DEFAULT
- 其他参数保持默认。

步骤5 创建完成后，即可在“Catalog”页面查看相关信息。

----结束

创建数据库

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择[创建LakeFormation实例及客户端连接](#)已创建的LakeFormation实例（例如lakeformation-test），在左侧导航栏选择“元数据 > 数据库”。

步骤4 在右上角“Catalog”后的下拉框中选择“hive”。

如果当前已包含名称为“default”的数据库，则跳过数据库的创建操作。

步骤5 单击“创建数据库”，配置相关参数后，单击“提交”。

- 库名称: default
固定名称，不可自定义。
- 所属Catalog: hive
- 选择位置: 单击“+”，选择位置，例如选择“obs://lakeformation-test/catalog1/database1”，单击“确定”。
- 其他参数保持默认。

步骤6 创建完成后，即可在“数据库”页面查看详细信息。

----结束

创建数据表

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入LakeFormation控制台。

步骤3 在左侧下拉框中选择[创建LakeFormation实例及客户端连接](#)已创建的LakeFormation实例（例如lakeformation-test），选择“元数据 > 表”，在右上角“Catalog”和“数据库”后的下拉框中分别选择“hive”、“default”。

步骤4 单击“创建表”，配置相关参数后，单击“提交”。

- 表名称：自定义，例如“table_A”。
- 数据存储位置：单击“+”，选择表存储在OBS并行文件系统中的位置，例如选择“obs://lakeformation-test/catalog1/database1/table1”，单击“确定”。
- 其他参数请根据实际需要进行配置。

步骤5 参考**步骤4**创建名称为“table_B”的数据表，存储路径为“lakeformation-test/catalog1/database1/table2”。

步骤6 创建完成后，即可在数据表页面查看相关信息。

----结束

为元数据授权

步骤1 登录管理控制台。

步骤2 在左上角单击“☰”，选择“大数据 > 湖仓构建 LakeFormation”进入 LakeFormation 控制台。

步骤3 在左侧下拉框中选择待操作的 LakeFormation 实例，选择“数据权限 > 数据授权”。

步骤4 单击“授权”，根据实际需要对hive数据目录、default数据库等进行授权，单击“确定”。

更多授权详细配置及要求请参考[新增授权](#)章节。

----结束

4 MRS 对接 LakeFormation 实例

MRS 对接前准备

MRS对接LakeFormation实例前准备请参考[准备工作](#)章节创建委托、配置LakeFormation数据连接等操作。

MRS 对接 LakeFormation 实例

当前仅MRS 3.3.0-LTS版本集群支持对接LakeFormation。

MRS对接LakeFormation实例，相关操作请参考[配置MRS 3.3.0-LTS版本集群对接LakeFormation](#)章节。