

LakeFormation

快速入门

文档版本 01
发布日期 2026-09-18



版权所有 © 华为云计算技术有限公司 2026。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 创建 LakeFormation 实例并规划元数据..... 1

2 为 LakeFormation 角色授予操作 Catalog 的权限..... 8

3 入门实践..... 13

1 创建 LakeFormation 实例并规划元数据

操作场景

在数据湖建设过程中，企业常面临元数据分散管理导致的数据孤岛问题，而传统方式难以对海量元数据进行统一管理。LakeFormation提供Catalog、数据库、数据表等的创建、修改、查看、删除等功能，通过集中式的统一管理帮助您便捷地进行数据湖初始化构建及持续运营，有效解决元数据分散管理问题，加速实现数据湖承载的业务规划和部署。

本入门提供从零开始创建LakeFormation实例，并为实例创建Catalog及内部的数据库、表等元数据的操作指导。

操作流程

开始使用如下样例前，请务必按[准备工作](#)指导完成必要操作。

1. [创建LakeFormation实例](#)：创建一个独享型LakeFormation实例。
2. [创建存储元数据的OBS路径](#)：创建用于存储元数据的OBS路径。
3. [创建Catalog](#)：创建一个名称为“catalog1”的Catalog。
4. [创建数据库](#)：在名称为“catalog1”的Catalog下创建一个名称为“database1”的数据库。
5. [创建数据表](#)：在名称为“database1”的数据库下创建一个名称为“table_A”的数据表。

准备工作

已准备用于创建并操作LakeFormation的用户。

- 注册账号并实名认证。
在创建LakeFormation实例之前，请先[注册华为账号并开通华为云](#)，进行[实名认证](#)。
如果您已开通华为云并进行实名认证，请忽略此步骤。
- 已准备具有创建LakeFormation实例权限的IAM用户，详细操作请参见[创建IAM用户并授权使用LakeFormation](#)。

步骤一：创建 LakeFormation 实例

- 步骤1** 使用已准备的用户登录[LakeFormation管理控制台](#)。
- 步骤2** 在“服务授权”页面，勾选“同意LakeFormation服务声明”，单击“同意授权”，完成服务授权。
- 如果已授权则跳过该步骤。
- 步骤3** 单击总览页面右上角“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 1-1 创建 LakeFormation 实例参数配置

参数	样例	参数说明
类型	独享	选择实例类型。 - 共享 - 独享
计费模式	按需收费	实例的计费模式。
项目	xxx	选择实例所属的项目。
名称	lakeformation-test	自定义LakeFormation实例名称。
QPS	10000	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。
企业项目	xxx	选择实例所属的企业项目。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。
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

步骤二：创建存储元数据的 OBS 路径

- 步骤1 登录[OBS管理控制台](#)。
- 步骤2 选择“并行文件系统 > 创建并行文件系统”，进入创建页面，配置相关参数后单击“立即创建”。
- 文件系统名称：根据界面要求设置并行文件系统名称，例如“lakeformation-test”。
 - 其他参数根据实际情况选择。
- 步骤3 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4 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/table2
 - 函数存储路径：lakeformation-test/catalog1/database1/udf1
- 结束

步骤三：创建 Catalog

- 步骤1 登录[LakeFormation管理控制台](#)。
- 步骤2 在左侧下拉框中选择已创建的LakeFormation实例（例如lakeformation-test），在左侧导航栏选择“数据目录”。
- 步骤3 单击“创建”，参考下表配置相关参数，其他参数保持默认，单击“提交”。

表 1-2 创建 Catalog

参数	样例	参数说明
Catalog名称	catalog1	填写待创建Catalog名称。 只能包含字母、数字和下划线，长度为1~256个字符。
Catalog类型	DEFAULT	选择Catalog类型。

参数	样例	参数说明
选择位置	obs://lakeformation-test/catalog1	Catalog数据存储在OBS中的位置。可选参数。 单击“+”，根据实际需要选择“并行文件系统”或“对象存储桶”，选择位置后，单击“确定”。 - 如果配置该参数，则所选位置只能以“obs://”开头，且必须包含一个存储对象，例如选择“obs://lakeformation-test/catalog1”。如果没有合适的OBS路径，可以单击“前往OBS创建”参考步骤二：创建存储元数据的OBS路径进行创建。 - 该路径不能与其他LakeFormation实例元数据存储路径重复，以免造成数据冲突。 - 建议选择未被其他Catalog选中的文件夹。
描述	xxx	所创建Catalog的描述信息。

步骤4 创建完成后，单击Catalog名称即可查看相关信息。

----结束

步骤四：创建数据库

- 步骤1** 登录[LakeFormation管理控制台](#)。
- 步骤2** 在左侧下拉框中选择已创建的LakeFormation实例（例如lakeformation-test），在左侧导航栏选择“数据目录”。
- 步骤3** 单击待创建数据库所属Catalog名称（例如“catalog1”），可以在“数据库”页签查看当前Catalog中包含的数据库。
- 步骤4** 在“数据库”页签单击“创建”，参考下表配置相关参数，其他参数保持默认，单击“提交”。

表 1-3 创建数据库

参数	样例	参数说明
库名称	database1	填写待创建数据库名称。 只能包含中文、字母、数字、下划线、中划线，长度为1~128个字符。
所属Catalog	catalog1	待创建数据库所属Catalog。

参数	样例	参数说明
选择位置	obs:// lakeformation -test/ catalog1/ database1	数据库元数据存储于OBS中的位置。 单击“+”，根据实际需要选择“并行文件系统”或“对象存储桶”，选择位置后，单击“确定”。 - 所选位置只能以“obs://”开头，且必须包含一个存储对象，例如选择“obs://lakeformation-test/catalog1/database1”。如果没有合适的OBS路径，可以单击“前往OBS创建”参考步骤二：创建存储元数据的OBS路径进行创建。 - 该路径必须与所属的Catalog存储路径（即创建Catalog时配置的“选择位置”参数）不同。 - 该路径不能与其他LakeFormation实例元数据存储路径重复，以免造成数据冲突。 - 如果所属Catalog配置了“数据库存储位置”参数，则此处该参数必须选择为所属Catalog“选择位置”的子路径、或“数据库存储位置”的子路径。
描述	xxx	所创建数据库的描述信息。

步骤5 创建完成后，即可在“数据库”页面查看详细信息。

----结束

步骤五：创建数据表

- 步骤1** 登录[LakeFormation管理控制台](#)。
- 步骤2** 在左侧下拉框中选择已创建的LakeFormation实例（例如lakeformation-test），选择“数据目录”。
- 步骤3** 展开待创建数据表所属Catalog名称（例如“catalog1”），单击待创建数据表所属数据库名称（例如“database1”），选择“数据表”页签。
- 步骤4** 在“数据表”页签单击“创建”，参考下表配置相关参数，其他参数保持默认，单击“提交”。

表 1-4 基本信息配置参数

参数	样例	参数说明
表名称	table_A	填写待创建的元数据表名称。 只能包含中文、字母、数字、下划线、中划线，长度为1~256个字符。
所属Catalog	catalog1	待创建表所属的Catalog。

参数	样例	参数说明
所属数据库	database1	待创建表所属的数据库。
表类型	MANAGED_TABLE	待创建表的类型，目前支持以下类型： - MANAGED_TABLE：管理表。如果删除管理表或分区，则与该表或分区关联的数据和元数据将被删除。 - EXTERNAL_TABLE：外部表。当文件已经存在或位于远程位置时，使用外部表。 - VIRTUAL_VIEW：虚拟视图。不存储实际的数据，不占用物理空间。 - MATERIALIZED_VIEW：物化视图。存储实际的数据，占用物理空间。
存储地址	obs://lakeformation-test/catalog1/database1/table1	表所映射的OBS的文件目录。 单击“”，选择表存储在OBS中的位置，单击“确定”。 - 可选参数，如果不配置，则数据表存储路径为“上层数据库存储路径/表名”。 - 如果配置该参数，则所选位置只能以“obs://”或“hdfs://”开头，且必须包含一个存储对象，例如选择“obs://lakeformation-test/catalog1/database1/table1”。如果没有合适的并行文件系统，可以单击“前往OBS创建”参考步骤二：创建存储元数据的OBS路径进行创建。 - 该路径必须与所属的Catalog、数据库的存储路径不同。 - 该路径不能与其他LakeFormation实例元数据存储路径重复，以免造成数据冲突。 - 如果所属数据库配置了“数据表存储位置”参数，则此处存储位置必须选择为所属数据库“选择位置”的子路径、或“数据表存储位置”的子路径。
描述	xxx	所创建表的描述信息。 长度为0~4000字节，1个中文字符对应3个字节。

参数	样例	参数说明
表格格式	HIVE	选择所创建表的表格格式。 ● HIVE：标准的表格格式。适用于传统数据仓库、离线批处理、Hadoop生态中的结构化数据查询。 ● ICEBERG：Apache Iceberg表格格式。适用于大规模结构化数据管理、实时分析、ACID事务需求高的场景。 ● LANCE：LANCE表格格式。适用于向量数据库、机器学习模型推理、高维数据相似性搜索。
数据源格式	Parquet	待创建表的数据源格式。
分隔符	-	数据源格式为“Csv”时，需设置字段分隔符。
Schema	-	配置表的Schema信息。 单击“Schema”区域的“添加”，根据实际需求配置相关参数。
表属性	-	配置表属性。 单击“表属性”区域的“添加”，并根据实际需要配置“Key”和“Value”的值。

步骤5 创建完成后，即可在数据表页面查看详细信息。

----结束

2 为 LakeFormation 角色授予操作 Catalog 的权限

操作场景

在企业数据治理场景中，当需要统一管理多源数据权限时，传统方式往往面临权限配置复杂、粒度粗的问题。通过LakeFormation服务，您可以创建实例，并针对该实例下的所有授权类型（Catalog、数据库、数据表、函数、OBS路径），授予授权主体（用户组、角色、IAM用户、委托用户）细粒度的数据访问权限。

本入门提供从零开始创建LakeFormation实例及其Catalog元数据，并创建LakeFormation角色，对角色授予修改Catalog、创建数据库的权限。授权后，绑定该角色的用户将同时拥有该角色的权限。

操作流程

开始使用如下样例前，请务必按[准备工作](#)指导完成必要操作。

1. [创建LakeFormation实例](#)：创建一个独享型LakeFormation实例。
2. [创建存储元数据的OBS路径](#)：创建用于存储元数据的OBS路径。
3. [创建Catalog](#)：创建一个名称为“catalog1”的Catalog。
4. [创建LakeFormation角色](#)：创建一个名为“lakeformation_role”的角色，并为当前用户绑定该角色。
5. [为角色授予操作Catalog的权限](#)：为“lakeformation_role”授予“catalog1”的修改、创建数据库的权限。

准备工作

已准备用于创建并操作LakeFormation的用户。

- 注册账号并实名认证。
在创建LakeFormation实例之前，请先[注册华为账号并开通华为云](#)，进行[实名认证](#)。
如果您已开通华为云并进行实名认证，请忽略此步骤。
- 已准备具有创建LakeFormation实例权限的IAM用户，详细操作请参见[创建IAM用户并授权使用LakeFormation](#)。

步骤一：创建 LakeFormation 实例

- 步骤1** 使用已准备的用户登录[LakeFormation管理控制台](#)。
- 步骤2** 在“服务授权”页面，勾选“同意LakeFormation服务声明”，单击“同意授权”，完成服务授权。
- 如果已授权则跳过该步骤。
- 步骤3** 单击总览页面右上角“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	样例	参数说明
类型	独享	选择实例类型。 - 共享 - 独享
计费模式	按需收费	实例的计费模式。
项目	xxx	选择实例所属的项目。
名称	lakeformation-test	自定义LakeFormation实例名称。
QPS	10000	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。
企业项目	xxx	选择实例所属的企业项目。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。
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

步骤二：创建存储元数据的 OBS 路径

- 步骤1** 登录[OBS管理控制台](#)。
- 步骤2** 选择“并行文件系统 > 创建并行文件系统”，进入创建页面，配置相关参数后单击“立即创建”。
- 文件系统名称：根据界面要求设置并行文件系统名称，例如“lakeformation-test”。
 - 其他参数根据实际情况选择。
- 步骤3** 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4** 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
- 结束

步骤三：创建 Catalog

- 步骤1** 登录[LakeFormation管理控制台](#)。
- 步骤2** 在左侧下拉框中选择已创建的LakeFormation实例（例如lakeformation-test），在左侧导航栏选择“数据目录”。
- 步骤3** 单击“创建”，参考下表配置相关参数，其他参数保持默认，单击“提交”。

表 2-2 创建 Catalog

参数	样例	参数说明
Catalog名称	catalog1	填写待创建Catalog名称。 只能包含字母、数字和下划线，长度为1~256个字符。
Catalog类型	DEFAULT	选择Catalog类型。
选择位置	obs://lakeformation-test/catalog1	Catalog数据存储在OBS中的位置。可选参数。 单击“+”，根据实际需要选择“并行文件系统”或“对象存储桶”，选择位置后，单击“确定”。 - 如果配置该参数，则所选位置只能以“obs://”开头，且必须包含一个存储对象，例如选择“obs://lakeformation-test/catalog1”。如果没有合适的OBS路径，可以单击“前往OBS创建”参考步骤二：创建存储元数据的OBS路径进行创建。 - 该路径不能与其他LakeFormation实例元数据存储路径重复，以免造成数据冲突。 - 建议选择未被其他Catalog选中的文件夹。

参数	样例	参数说明
描述	xxx	所创建Catalog的描述信息。

步骤4 创建完成后，单击Catalog名称即可查看相关信息。

----结束

步骤四：创建 LakeFormation 角色

- 步骤1** 登录[LakeFormation管理控制台](#)。
- 步骤2** 在左侧下拉框中选择待操作的LakeFormation实例，选择“数据权限 > 角色”。
- 步骤3** 单击“创建角色”，填写角色名称，例如“lakeformation_role”，单击“确定”。
- 步骤4** 在新创建的角色所在行，单击操作列的“添加IAM用户”，在下拉框中选择待绑定的用户，例如选择[准备工作](#)中准备的用户，单击“确定”。
- 结束

步骤五：为角色授予操作 Catalog 的权限

- 步骤1** 登录[LakeFormation管理控制台](#)。
- 步骤2** 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3** 单击“授权”，在弹出的窗口中参考下表配置参数后，单击“确定”。

表 2-3 元数据权限授权

参数	样例	参数说明
主体类型	角色	待授权的主体类型。取值范围： • 用户组 • 角色 • IAM用户 • 委托用户
选择角色	lakeformation_role	选择待授权的主体名称。名称中不能包含中划线（-），否则可能造成操作失败。 该参数名称与选择的“主体类型”有关。
授权类型	资源	• 资源：表示对LakeFormation实例中的资源进行授权。 • 路径：表示对OBS路径进行授权。
Catalog	catalog1	选择待授权的“Catalog”。

参数	样例	参数说明
操作类型	ALTER、 CREATE_DATA BASE	选择授权的操作类型。不同的授权类型对应的操作类型不同。Catalog支持的操作类型如下： • ALL：Catalog的所有操作权限。 • ALTER：修改Catalog。 • CREATE_DATABASE：创建数据库。 • DROP：删除Catalog。 • DESCRIBE：查看Catalog的元数据信息或切换Catalog。 • LIST_DATABASE：查看Catalog下资源列表。
赋予授权权限	-	是否授予授权权限。 赋予授权权限后，授权主体便拥有将对象授权其他授权主体的权限。

----结束

3 入门实践

当用户完成LakeFormation实例创建后，可以根据自身的业务需求使用LakeFormation提供的一系列常用实践。

表 3-1 常用最佳实践

实践		描述
与开源组件对接使用	配置LakeFormation对接开源Spark组件	本实践指导用户对接开源Spark组件的操作，并提供以下二次开发样例。 - 自定义认证信息获取类：用于获取访问LakeFormation服务的IAM认证信息。 - 自定义用户信息获取类：用于获取当前访问LakeFormation的用户。
	配置LakeFormation对接开源Hive组件	本实践指导用户对接开源Hive组件的操作，并提供以下二次开发样例。 - 自定义认证信息获取类：用于获取访问LakeFormation服务的IAM认证信息。 - 自定义用户信息获取类：用于获取当前访问LakeFormation的用户。