LakeFormation

产品介绍

文档版本 01

发布日期 2025-01-10

华为云计算技术有限公司

商标声明


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 什么是 LakeFormation

湖仓构建（DataArts Lake Formation，简称LakeFormation）是企业级数据湖一站式构建服务，在存算分离架构基础上提供数据湖元数据统一管理的可视化界面及API，兼容Hive元数据模型以及Ranger权限模型，支持对接MapReduce服务（MRS）、数据仓库服务 GaussDB（DWS）、数据湖探索（DLI）、AI开发平台ModelArts、数据治理中心 DataArts Studio等多种计算引擎及大数据云服务，使用户可以便捷高效地构建数据湖和运营相关业务，加速释放业务数据价值。

LakeFormation产品通过底层资源实现跨AZ部署及高可靠、弹性伸缩、元数据统一管理、元数据与文件目录联动授权、对接多计算引擎等功能，是一个Serverless服务。

LakeFormation 架构

LakeFormation服务架构图如图1-1所示。

图 1-1 LakeFormation 服务架构

LakeFormation功能包括元数据管理、数据权限管理、控制台、API。

- 元数据基于Hive元数据模型，支持Catalog、数据库、表、函数等元数据对象。
- 数据权限管理提供权限策略的配置和对应的权限访问控制。
  - 授权主体支持IAM用户和用户组以及LakeFormation角色。
  - 授权对象支持Catalog、数据库、表及列、函数等元数据对象，也支持OBS并行文件系统路径。
  - 授权操作包含元数据对象的相关操作，以及OBS路径的读写操作。
- Console支持实例管理、元数据管理、数据权限管理、接入管理、任务管理等操作。
- API层提供支持兼容Hive社区的元数据接口，以及兼容Ranger社区的权限同步接口，以便于MRS、DWS等服务的集成对接。

产品优势

- 生态开放

  遵循开源事实标准，支撑存量业务平滑演进。
  - 提供兼容Hive/Spark/Flink/Trino社区的元数据接口，支持计算引擎平滑对接。
  - 提供兼容Ranger的权限接口，一次授权，统一生效。
  - 提供迁移工具，支持存量MRS集群相关元数据的平滑迁移。
- 数智融合

  打通大数据的数据壁垒，实现真正数智融合。
  - 支持数据库、表、函数、模型、非结构化数据集等统一管理。
  - 实现统一的细粒度数据权限管理，支持跨服务/跨集群的数据共享。
- 大规格高可靠

  支撑超大规模大数据业务的高可靠。
  - 超大规模元数据管理能力。
  - 统一权限管理能力，支持海量细粒度权限管理。
  - 支持多AZ的容灾能力。
- 简单易用

  提供基于元数据的增值管理能力。
  - Serverless架构，开箱即用。
  - 提供数据湖管理、元数据统计等管理能力。

产品功能

表1-1列出了湖仓构建LakeFormation提供的常用功能特性。

在使用LakeFormation之前，建议您先了解湖仓构建服务LakeFormation的基本概念，以便更好地理解LakeFormation提供的各项功能。

表 1-1 湖仓构建服务 LakeFormation 功能概览

功能名称	功能描述
实例类型	LakeFormation提供了不同实例类别，满足不同场景下客户对性能和成本的不同诉求。具体介绍请参考**共享型与独享型LakeFormation对比**。
实例管理	LakeFormation提供实例的创建、总览、删除等基本功能，帮助您便捷地进行实例管理，加速实现数据湖承载的业务的规划和部署。
元数据管理	LakeFormation提供数据湖元数据Catalog、数据库、数据表等的创建、修改、查看、删除等功能，并支持配置元数据生命周期。帮助您便捷地进行数据湖初始化构建以及持续运营，集中式的统一管理LakeFormation实例下所有的元数据，加速实现数据湖承载的业务的规划和部署。
数据权限管理	LakeFormation提供针对Catalog、Database、Table等数据资源的授权、取消、查看等功能。帮助您对数据湖实现便捷的统一的数据权限管理。
任务管理	LakeFormation支持将外部服务的元数据及其权限全量或增量迁移至当前LakeFormation实例中，对元数据及权限进行统一管理。
接入管理	LakeFormation提供统一的接入管理能力，用户可以通过创建接入客户端的方式为指定的客户端环境建立网络连接通道，同时可以在客户端详情中查看接入IP、接入域名等信息，用于其他云服务接入LakeFormation实例。

访问方式

当前提供了Web化的服务管理平台，即管理控制台和基于HTTPS请求的API（Application programming interface）管理方式。除此外，LakeFormation也提供SDK客户端，更进一步方便计算引擎的对接集成。

- API方式

  如果用户需要将公有云平台上的LakeFormation实例集成到第三方系统，用于二次开发，可使用API方式访问LakeFormation实例，具体操作请参见**API参考**。
- 控制台方式

  如果用户已注册公有云，用户可使用管理控制台方式，从服务列表中选择“大数据 > 湖仓构建 LakeFormation”访问LakeFormation。
- SDK方式
  - LakeFormation提供兼容Hive元数据模型的SDK客户端，如果用户需要将Hive、Spark等计算引擎对接LakeFormation，用于统一元数据管理，可使用SDK方式访问LakeFormation实例。
  - LakeFormation提供了REST（Representational State Transfer）风格API，支持通过HTTPS请求调用。
  - SDK使用的具体操作请参见**SDK参考**。

2 应用场景

2.1 数据湖建设和持续运营

场景描述

数据湖建设和持续运营，是指数据湖的初始化建设及后续日常的海量元数据及权限管理，因此用户需要便捷高效的建设和管理方式。

传统方式的弊端

- 仅支持通过计算引擎（Hive、Spark等）执行SQL实现元数据的定义、修改、查询，对用户有一定的技能要求，缺少提升易用性的可视化界面。
- 一个完整的授权活动，需要针对计算引擎、对象存储执行两次授权操作，对用户操作带来不便，易用性差。

LakeFormation 服务优势

- 一站式可视化湖仓构建：提供数据湖元数据统一定义和授权的可视化界面，支持用户便捷操作，快速构建。
- 联动授权：支持在元数据授权的同时，自动化完成元数据所映射的文件目录的授权，使用户授权操作更便捷和高效。
- 细粒度访问控制：实现针对数据湖的库、表、列级元数据的细粒度访问控制，为业务数据的安全性提供有力保障。

建议搭配服务

MapReduce服务（MRS）

数据仓库服务 GaussDB（DWS）

数据治理中心 DataArts Studio

数据湖探索（DLI）

说明

支持情况请咨询对应服务。

2.2 多服务/多集群共享元数据

场景描述

多服务/多集群均使用统一的元数据，最大化实现数据的共享，避免不必要的重复数据，更大程度释放业务数据价值。

本服务的优势

- 兼容Hive元数据模型：提供兼容Hive元数据模型的SDK客户端，使计算引擎对接LakeFormation更轻松和高效。
- 兼容Ranger权限模型：提供兼容Ranger权限模型的接口，具备良好的生态扩展性。

建议搭配服务

MapReduce服务（ MRS ）

数据仓库服务 GaussDB（ DWS ）

数据湖探索（ DLI ）

说明

支持情况请咨询对应服务。

3 共享型与独享型 LakeFormation 对比

LakeFormation提供了不同实例类别，满足不同场景下用户对性能和成本的不同诉求。

说明

独享型特性当前仅针对白名单用户开放。

- 计费对比

  共享型与独享型实例的计费项及计费说明详细介绍请参考**计费说明**章节中**表4-1**。

- 性能对比

  表 3-1 性能对比

类型	共享型实例	独享型实例
部署模式	物理资源共享，实例间逻辑隔离。	物理资源独占，实例的性能不受其他实例的影响，可根据业务需要选择不同规格的实例。
每秒请求数（QPS）	保证2000每秒请求数。	按照用户创建实例时的选择不同。

- 功能对比

  表 3-2 功能对比

类型	描述	共享型实例	独享型实例
Catalog管理	LakeFormation提供数据湖中Catalog的元数据创建、修改、删除、查看等功能。	√	√
数据库管理	LakeFormation提供数据湖中数据库的元数据创建、修改、删除、查看等功能。	√	√

类型	描述	共享型实例	独享型实例
表管理	LakeFormation提供数据湖中数据表的元数据创建、修改、删除、查看等功能。	√	√
函数管理	LakeFormation提供数据湖中函数的元数据创建、修改、删除、查看等功能。	√	√
元数据生命周期管理	LakeFormation支持配置数据的删除策略，节省空间及成本，提升系统的灵活性。	√	√
元数据权限管理	提供针对元数据的授权、取消、查看等功能。	√	√
元数据迁移管理	支持将外部服务的元数据全量或增量迁移至当前LakeFormation实例中，对元数据进行统一管理。	√	√
权限迁移管理	支持将外部服务的元数据权限全量或增量迁移至当前LakeFormation实例中，对元数据的权限进行统一管理。	√	√
接入客户端管理	提供统一的接入管理能力，用户可以通过创建接入客户端的方式为指定的客户端环境建立网络连接通道，用于其他云服务接入LakeFormation实例。	√	√

4 计费说明

计费项

华为云湖仓构建服务LakeFormation根据您选择的实例规格和使用时长计费。

详细的计费项及说明请参考表4-1。

您也可以通过LakeFormation提供的价格计算器，选择您需要的实例规格和使用时长等，来快速计算出购买LakeFormation实例的参考价格。

表 4-1 计费项信息

实例规格	计费项	计费说明
独享型	元数据对象数量	按照元数据对象使用量收费，元数据对象数量为Catalog、数据库、表、分区、索引、函数数量之和。 按照万个/小时计费，不足万个按照万个计算。
	每秒查询率（QPS）	按照用户购买时选择的QPS上限进行计费。当前支持购买1W至5W五种QPS规格。
共享型	元数据对象数量	按照元数据对象使用量收费，元数据对象数量为Catalog、数据库、表、分区、索引、函数数量之和。 前100万个不计费，后续按照万个/小时计费，不足万个按照万个计算。
	API调用次数	按照元数据相关API的调用次数收费，每个月前100万次不收费，后续按每次计费。

计费模式

LakeFormation当前支持按需计费模式。详细的计费说明请参考表4-2。

表 4-2 计费模式

计费模式	按需计费

付费方式	后付费 按照LakeFormation实例实际使用时长计费。
计费周期	秒级计费，按小时结算。
更改计费模式	暂不支持。
变更规格	支持变更实例规格。
适用场景	适用于使用需求波动的场景，可以随时开通，随时删除。

到期与欠费

用户欠费后，可以查看欠费详情。为防止相关资源被停止或者释放，需要用户及时进行充值。如果账户余额不足，账号将进入欠费状态，需要在约定时间内支付欠款，详细操作请参考**欠费还款**。

如果没有及时地进行续费或充值，将进入宽限期。如宽限期满仍未续费或充值，将进入保留期。在保留期内资源将停止服务。保留期满仍未续费或充值，存储在云服务中的数据将被删除、云服务资源将被释放。详细说明请参考“**资源停止服务或逾期释放说明**”。宽限期与保留期的具体规则请参考“**宽限期保留期**”。

5 基本概念

5.1 元数据

数据目录（Catalog）

LakeFormation实例的元数据资源中的最顶层资源，即在一个LakeFormation实例下可以创建多个Catalog，包含名称、描述、位置等信息，支持创建、修改、删除等操作。

其中位置是Catalog所映射的OBS并行文件系统的文件目录。

数据库（Database）

LakeFormation实例的数据目录（Catalog）的下级资源，即在一个Catalog下可以创建多个Database，包含名称、所属Catalog、拥有者、位置、描述等信息，支持创建、修改、删除以及授权和查看权限等操作。

其中位置是Database所映射的OBS并行文件系统的文件目录。

数据表（Table）

LakeFormation实例的数据库（Database）的下级资源，即在一个Database下可以创建多个Table，包含基本信息、格式与序列化信息、字段信息、属性信息，支持创建、修改、删除以及授权和查看权限等操作。

函数（Function）

在SQL查询中使用函数对数据进行特定处理，包括内置函数和用户自定义函数UDF（User-Defined Functions）。

用户自定义函数分为以下几类：

- 普通的UDF，用于操作单个数据行，且产生一个数据行作为输出。
- 用户定义聚集函数UDAF（User-Defined Aggregating Functions），用于接受多个输入数据行，并产生一个输出数据行。
- 用户定义表生成函数UDTF（User-Defined Table-Generating Functions），用于操作单个输入行，产生多个输出行。

分区（Partition）

分区是对数据表按照行维度进行分割，目的是为了在特定SQL操作中减少数据读写的总量以缩减响应时间。

5.2 数据权限

权限策略

用户可以在管理控制台的LakeFormation实例界面，针对该实例下的所有Catalog、Database、Table等数据资源，授予用户组等主体细粒度的数据访问权限。

经过以上授权操作，形成一条或多条权限策略。

权限策略包含授权主体、授权对象、权限、授权权限，支持取消本条权限策略的操作。

授权主体

使其具备针对某数据资源的指定访问权限的用户/用户组/角色等身份，如某一用户组、某一角色等。

授权主体类型包括“GROUP”、“ROLE”、“USER”等。

- 用户（USER）： 华为云IAM用户
- 用户组（GROUP）： 华为云IAM用户组
- 角色（ROLE）： LakeFormation角色

授权对象

LakeFormation中管理的元数据对象，包含Catalog、Database、Table等数据资源，如某一数据库、某些数据表的列等。允许授权的资源类型包括“CATALOG”、“DATABASE”、“TABLE”、“COLUMN”、“FUNC”等。

- 数据目录（CATALOG）： LakeFormation管理的数据目录，可以包含多个数据库。
- 数据库（DATABASE）： LakeFormation管理的数据库，可以包含多个数据表或函数。
- 数据表（TABLE）： LakeFormation管理的数据表，可以包含多个列。
- 列（COLUMN）： LakeFormation管理的列。
- 函数（FUNC）： LakeFormation管理的函数。

权限

使用户具备针对某数据资源的具体访问/操作权限，如“ALTER”、“DROP”、“ALL”等。每种资源允许被授予的权限请参考**表6-4**。

授权权限

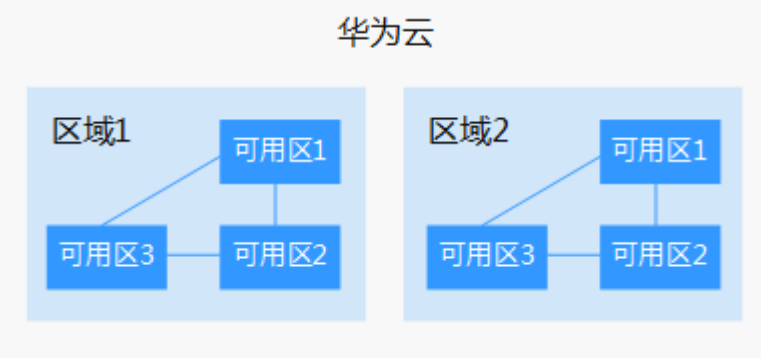
在用户已具备针对某数据资源的某些访问/操作权限的前提下，“授权权限”使该用户能够将已具备权限再次授予给其他用户。

5.3 区域与可用区

通常用区域和可用区来描述数据中心的位置，您可以在特定的区域、可用区创建资源。

- 区域（Region）指物理的数据中心。每个区域完全独立，这样可以实现一定程度的容错能力和稳定性。资源创建成功后不能更换区域。
- 可用区（AZ，Availability Zone）是同一区域内，电力和网络互相隔离的物理区域，一个可用区不受其他可用区故障的影响。一个区域内可以有多个可用区，不同可用区之间物理隔离，但内网互通，既保障了可用区的独立性，又提供了低价、低时延的网络连接。

图 5-1 区域和可用区



目前，华为云已在全球多个地域开放云服务，您可以根据需求选择适合自己的区域和可用区。更多信息请参见**华为云全球站点**。

6 权限管理

6.1 LakeFormation 权限概述

LakeFormation对于元数据权限和数据权限的管理，使用“粗粒度”的Identity and Access Management（简称IAM）权限和“细粒度”的LakeFormation权限的组合，实现精细访问控制。

- “粗粒度”的IAM权限：对于各项操作有更广泛的权限。例如，推荐用户使用“lakeformation:*:create”（即LakeFormation所有元数据的创建权限），而不是使用“lakeformation:table:create”（即LakeFormation数据表的创建权限）来控制用户对于表的创建权限，同时使用“细粒度”LakeFormation权限“CREATE_TABLE”来控制用户是否能够在某个Database下创建一个Table元数据。
- “细粒度”的LakeFormation权限：指使用LakeFormation权限，向各个主体（包括用户、用户组、角色等）授予对于元数据、OBS路径以及其中的数据的访问权限。

IAM权限模型由IAM策略组成。LakeFormation权限模型使用LakeFormation定义的权限主体、授权对象、权限组成，详细介绍请参考**基本概念**。

当用户请求访问元数据或数据时，请求必须通过IAM和LakeFormation的权限检查才能成功。

6.2 IAM 权限介绍

IAM 概述

如果您需要对华为云上购买的LakeFormation资源，为企业中的员工设置不同的访问权限，以达到不同员工之间的权限隔离，您可以使用统一身份认证服务（Identity and Access Management，简称IAM）进行精细的权限管理。该服务提供用户身份认证、权限分配、访问控制等功能，可以帮助您安全的控制华为云资源的访问。

通过IAM，您可以在华为云账号中给员工创建IAM用户，并授权控制其对华为云资源的访问范围。例如您希望其拥有LakeFormation的使用权限，但是不希望其拥有删除数据库等高危操作的权限，那么您可以使用IAM创建用户，通过授予仅能查询LakeFormation实例，但是不允许删除的权限，控制其对云服务资源的使用范围。

如果华为云账号已经能满足您的要求，不需要创建独立的IAM用户进行权限管理，您可以跳过对应权限管理操作。

IAM是华为云提供权限管理的基础服务，无需付费即可使用，您只需要为您账号中的资源进行付费。关于IAM的详细介绍，请参见**IAM产品介绍**。

IAM 权限

默认情况下，管理员创建的IAM用户没有任何权限，需要将其加入用户组，并给用户组授予策略或角色，才能使得用户组中的用户获得对应的权限，这一过程称为授权。授权后，用户就可以基于被授予的权限对云服务进行操作。

LakeFormation授权时，在全局级服务中设置权限，不需要切换区域。

权限根据授权精细程度分为角色和策略。

- 角色：IAM最初提供的一种根据用户的工作职能定义权限的粗粒度授权机制。该机制以服务为粒度，提供有限的服务相关角色用于授权。由于华为云各服务之间存在业务依赖关系，因此给用户授予角色时，可能需要一并授予依赖的其他角色，才能正确完成业务。角色并不能满足用户对精细化授权的要求，无法完全达到企业对权限最小化的安全管控要求。
- 策略：IAM最新提供的一种细粒度授权的能力，可以精确到具体服务的操作、资源以及请求条件等。基于策略的授权是一种更加灵活的授权方式，能够满足企业对权限最小化的安全管控要求。例如：针对LakeFormation服务，管理员能够控制IAM用户仅能对某一类云服务器资源进行指定的管理操作。

IAM 系统策略

表6-1介绍了LakeFormation的默认系统策略。

表 6-1 LakeFormation 系统策略

系统角色/策略名称	描述	类别	依赖关系
LakeFormation FullAccess	LakeFormation管理员权限，拥有该权限的用户可以操作并使用所有LakeFormation服务功能。	系统策略	• IAM AgencyFullAccess • OBS OperateAccess • VPC FullAccess • VPCEndpoint FullAccess

系统角色/策略名称	描述	类别	依赖关系
LakeFormation ReadOnlyAccess	LakeFormation只读权限，拥有该权限的用户可以执行LakeFormation所有查询类功能。	系统策略	• IAM ReadOnlyAccess • OBS ReadOnlyAccess • VPC ReadOnlyAccess • VPCEndpoint ReadOnlyAccess
LakeFormation CommonOperations	LakeFormation基础权限，包含LakeFormation服务协议查看/授权/取消，以及OBS、TMS等周边依赖服务的基础权限集合。	系统策略	• IAM ReadOnlyAccess • OBS ReadOnlyAccess • VPC FullAccess • VPCEndpoint FullAccess

表6-2介绍了lakeFormation常用操作与系统权限的授权关系，您可以参考该表选择合适的系统权限。

表 6-2 LakeFormation 常用操作与系统策略的授权关系

操作	LakeFormation FullAccess	LakeFormation CommonOperations	LakeFormation ReadOnlyAccess
查询LakeFormation实例	√	√	√
创建LakeFormation实例	√	x	x
变更LakeFormation实例	√	√	x
删除LakeFormation实例	√	x	x

操作	LakeFormation FullAccess	LakeFormation CommonOperations	LakeFormation ReadOnlyAccess
恢复LakeFormation实例	√	x	x
查询租户所有标签	√	√	√
更新LakeFormation实例的标签	√	√	x
创建元数据迁移/发现任务	√	√	x
修改元数据迁移/发现任务	√	√	x
删除元数据迁移/发现任务	√	√	x
查询元数据迁移/发现任务	√	√	√
查询元数据迁移/发现任务日志	√	√	√
运行/停止元数据迁移/发现任务	√	√	x
同意用户协议	√	√	√
查询用户协议	√	√	√
删除用户协议	√	√	√
授权LakeFormation服务创建委托	√	x	x
查询LakeFormation服务创建的委托	√	√	√
删除LakeFormation服务创建的委托	√	x	x
授权资源	√	x	x
查询授权资源	√	√	√
取消授权资源	√	x	x
查询OBS桶列表	√	√	√

操作	LakeFormation FullAccess	LakeFormation CommonOperations	LakeFormation ReadOnlyAccess
查询OBS桶对象列表	√	√	√
创建服务接入客户端	√	√	x
查询服务接入客户端	√	√	√
删除服务接入客户端	√	√	x
订阅元数据事件	√	√	x
取消订阅元数据事件	√	√	x
查询元数据事件	√	√	√
查询Catalog元数据	√	√	√
创建Catalog元数据	√	√	x
修改Catalog元数据	√	√	x
删除Catalog元数据	√	√	x
查询Database元数据	√	√	√
创建Database元数据	√	√	x
修改Database元数据	√	√	x
删除Database元数据	√	√	x
查询Table元数据	√	√	√
创建Table元数据	√	√	x
修改Table元数据	√	√	x
删除Table元数据	√	√	x
查询Partition元数据	√	√	√

操作	LakeFormation FullAccess	LakeFormation CommonOperations	LakeFormation ReadOnlyAccess
创建Partition元数据	√	√	x
修改Partition元数据	√	√	x
删除Partition元数据	√	√	x
查询列统计信息	√	√	√
创建列统计信息	√	√	x
修改列统计信息	√	√	x
删除列统计信息	√	√	x
查询Function元数据	√	√	√
创建Function元数据	√	√	x
修改Function元数据	√	√	x
删除Function元数据	√	√	x
查询Model元数据	√	√	√
创建Model元数据	√	√	x
修改Model元数据	√	√	x
删除Model元数据	√	√	x
查询ModelFile元数据	√	√	√
创建ModelFile元数据	√	√	x
修改ModelFile元数据	√	√	x
删除ModelFile元数据	√	√	x
查询dataset元数据	√	√	√
创建dataset元数据	√	√	x
修改dataset元数据	√	√	x

操作	LakeFormation FullAccess	LakeFormation CommonOperations	LakeFormation ReadOnlyAccess
删除dataset元数据	√	√	x
查询元数据数量	√	√	√
查询授权主体	√	√	√
创建角色	√	√	x
删除角色	√	√	x
修改角色	√	√	x
查询角色	√	√	√
将用户/用户组加入角色	√	√	x
将用户/用户组移除角色	√	√	x
更新角色中的用户/用户组	√	√	x
将元数据权限授权给授权主体	√	√	x
取消授权元数据权限给授权主体	√	√	x
查询授权信息	√	√	√
获取访问数据的STSToken	√	√	x

LakeFormation系统策略所包含的详细内容如下：

- **LakeFormation FullAccess策略内容**

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Action": [
                "lakeformation:*:*",
                "vpc:*:get",
                "vpc:*:list",
                "tms:predefineTags:list",
                "obs:bucket:ListAllMyBuckets",
                "obs:bucket:ListBucket",
                "obs:bucket:HeadBucket",
                "obs:object:GetObject"
            ],
            "Effect": "Allow"
        }
    ]
}
```

- **LakeFormation CommonOperations策略**

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:*:describe*",
                "lakeformation:*:list*",
                "lakeformation:policy:export",
                "lakeformation:access:create",
                "lakeformation:access:delete",
                "lakeformation:accessAgency:describe",
                "lakeformation:accessService:describe",
                "lakeformation:accessService:grant",
                "lakeformation:accessTenant:grant",
                "lakeformation:agreement:cancel",
                "lakeformation:agreement:describe",
                "lakeformation:agreement:grant",
                "lakeformation:catalog:alter",
                "lakeformation:catalog:create",
                "lakeformation:catalog:drop",
                "lakeformation:database:alter",
                "lakeformation:database:create",
                "lakeformation:database:drop",
                "lakeformation:dataset:alter",
                "lakeformation:dataset:alterFile",
                "lakeformation:dataset:alterFileGroup",
                "lakeformation:dataset:create",
                "lakeformation:dataset:createFile",
                "lakeformation:dataset:createFileGroup",
                "lakeformation:dataset:drop",
                "lakeformation:dataset:dropFile",
                "lakeformation:dataset:dropFileGroup",
                "lakeformation:function:alter",
                "lakeformation:function:create",
                "lakeformation:function:drop",
                "lakeformation:group:alter",
                "lakeformation:instance:access",
                "lakeformation:instance:alter",
                "lakeformation:instanceJob:alter",
                "lakeformation:instanceJob:create",
                "lakeformation:instanceJob:drop",
                "lakeformation:instanceJob:exec",
                "lakeformation:job:alter",
                "lakeformation:job:create",
                "lakeformation:job:drop",
                "lakeformation:job:exec",
                "lakeformation:model:alter",
                "lakeformation:model:alterFile",
                "lakeformation:model:create",
                "lakeformation:model:createFile",
                "lakeformation:model:drop",
                "lakeformation:model:dropFile",
                "lakeformation:policy:create",
                "lakeformation:policy:drop",
                "lakeformation:role:alter",
                "lakeformation:role:create",
                "lakeformation:role:drop",
                "lakeformation:table:alter",
                "lakeformation:table:create",
                "lakeformation:table:drop",
                "lakeformation:transaction:operate",
                "lakeformation:user:alter",
                "vpc:*:get",
                "vpc:*:list",
                "tms:predefineTags:list",
                "obs:bucket:ListAllMyBuckets",
                "obs:bucket:ListBucket",
```

```
            "obs:bucket:HeadBucket",
            "obs:object:GetObject"
        ]
    }
  ]
}
```

- **LakeFormation ReadOnlyAccess策略**

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "lakeformation:*:describe*",
        "lakeformation:*:list*",
        "lakeformation:policy:export",
                "lakeformation:agreement:cancel",
                "lakeformation:agreement:describe",
                "lakeformation:agreement:grant",
        "vpc:*:get",
        "vpc:*:list",
        "tms:predefineTags:list",
        "obs:bucket:ListAllMyBuckets",
        "obs:bucket:ListBucket",
        "obs:bucket:HeadBucket",
        "obs:object:GetObject"
      ],
      "Effect": "Allow"
    }
  ]
}
```

LakeFormation 的 IAM 权限列表

表6-3列举了LakeFormation的所有IAM权限。

表 6-3 LakeFormation 的 IAM 权限列表

操作类型	操作项	描述
只读	lakeformation:access:describe	查询接入客户端。
	lakeformation:agency:describe	查询委托。
	lakeformation:catalog:describe	查询Catalog元数据。
	lakeformation:configuration:describe	查询配置。
	lakeformation:credential:describe	查询认证信息。
	lakeformation:database:describe	查询数据库元数据。
	lakeformation:file:describe	查询文件。
	lakeformation:function:describe	查询函数元数据。
	lakeformation:group:describe	查询用户组以及关联角色关系。

操作类型	操作项	描述
	lakeformation:instance:describe	查询实例。
	lakeformation:instance:listAuthorizedLocation	查询已授权给LakeFormation服务的OBS路径。
	lakeformation:instanceJob:describe	查询实例级任务。
	lakeformation:job:describe	查询任务。
	lakeformation:metadataEvent:describe	查询元数据事件。
	lakeformation:obs:describe	查询OBS桶列表。
	lakeformation:part:describe	查询分区。
	lakeformation:policy:describe	查询权限策略。
	lakeformation:policy:export	批量查询权限策略。
	lakeformation:role:describe	查询角色。
	lakeformation:table:describe	查询表元数据。
	lakeformation:tableFile:describe	查询文件。
	lakeformation:tableFileGroup:describe	查询表文件组元数据。
	lakeformation:tag:describe	查询资源标签。
	lakeformation:user:describe	查询用户以及关联角色关系。
写	lakeformation:access:create	创建接入客户端。
	lakeformation:access:delete	删除接入客户端。
	lakeformation:agency:create	创建委托。
	lakeformation:agency:drop	删除委托。
	lakeformation:catalog:alter	修改Catalog元数据。
	lakeformation:catalog:create	创建Catalog元数据。
	lakeformation:catalog:drop	删除Catalog元数据。
	lakeformation:database:alter	修改数据库元数据。
	lakeformation:database:create	创建数据库元数据。
	lakeformation:database:drop	删除数据库元数据。
	lakeformation:dataset:create	创建数据集元数据。

操作类型	操作项	描述
	lakeformation:file:create	创建文件。
	lakeformation:file:drop	删除文件。
	lakeformation:file:alter	修改文件。
	lakeformation:function:alter	修改函数元数据。
	lakeformation:function:create	创建函数元数据
	lakeformation:function:drop	删除函数元数据。
	lakeformation:group:alter	修改用户组以及关联角色关系。
	lakeformation:instance:access	申请接入服务。
	lakeformation:instance:alter	修改实例。
	lakeformation:instance:create	创建实例。
	lakeformation:instance:drop	删除实例。
	lakeformation:instanceJob:alter	修改任务。
	lakeformation:instanceJob:create	创建任务。
	lakeformation:instanceJob:drop	删除任务。
	lakeformation:instanceJob:exec	执行实例级任务。
	lakeformation:instance:createSubscriber	创建元数据事件订阅者。
	lakeformation:instance:deleteSubscriber	删除元数据事件订阅者。
	lakeformation:job:alter	修改任务。
	lakeformation:job:create	创建任务。
	lakeformation:job:drop	删除任务。
	lakeformation:job:exec	执行任务。
	lakeformation:model:create	创建模型元数据。
	lakeformation:metadata:restore	恢复元数据。
	lakeformation:part:alter	修改分区。
	lakeformation:part:drop	删除分区。
	lakeformation:part:create	创建分区。

操作类型	操作项	描述
	lakeformation:policy:create	创建权限策略。
	lakeformation:policy:delegate	将权限策略委托给其他授权主体。
	lakeformation:policy:drop	删除权限策略。
	lakeformation:role:alter	修改角色以及关联用户组关系。
	lakeformation:role:create	创建角色。
	lakeformation:role:drop	删除角色。
	lakeformation:table:alter	修改表元数据。
	lakeformation:table:create	创建表元数据。
	lakeformation:table:drop	删除表元数据。
	lakeformation:tableFile:alter	修改表文件。
	lakeformation:tableFile:create	创建表文件。
	lakeformation:tableFile:drop	删除表文件。
	lakeformation:tableFileGroup:alter	修改表文件组元数据。
	lakeformation:tableFileGroup:create	创建表文件组元数据。
	lakeformation:tableFileGroup:drop	删除表文件组元数据。
	lakeformation:transaction:operate	操作事务。
	lakeformation:user:alter	修改用户以及关联角色关系。
权限管理	lakeformation:accessService:grant	授权接入服务。
	lakeformation:accessTenant:grant	授权接入租户。
	lakeformation:accessAgency:describe	查询接入委托信息。
	lakeformation:accessService:describe	查看接入服务。
	lakeformation:agreement:describe	查询服务协议授权。
	lakeformation:agreement:cancel	取消服务协议授权。

操作类型	操作项	描述
	lakeformation:agreement:grant	授权服务协议授权。
	lakeformation:instance:authorizeLocation	授权将OBS路径授权给LakeFormation服务。
	lakeformation:instance:cancelAuthorizeLocation	取消授权OBS路径。

6.3 LakeFormation 权限介绍

LakeFormation 权限

用户可以在管理控制台的LakeFormation实例界面，针对该实例下的所有Catalog、Database、Table等数据资源，授予用户组等主体细粒度的数据访问权限。

经过以上授权操作，形成一条或多条权限策略，权限策略包含授权主体、授权对象、权限、授权权限等。

表6-4介绍了不同元数据类型的LakeFormation权限：

表 6-4 不同授权类型的操作权限

授权类型	操作类型	权限说明
Catalog	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。
数据库	ALL	数据库的所有操作权限。
	ALTER	修改数据库。
	DROP	删除数据库。
	DESCRIBE	查看数据库的元数据信息或切换数据库。
	LIST_TABLE	查看数据库下资源列表。
	LIST_FUNC	查看某一数据库下的函数。
	CREATE_TABLE	在数据库中创建表。

授权类型	操作类型	权限说明
	CREATE_FUNC	在数据库中创建函数。
表	ALL	表的所有操作权限。
	ALTER	修改表。
	DROP	删除表。
	DESCRIBE	查看表的元数据信息。
	UPDATE	更新表数据。
	INSERT	插入表数据。
	SELECT	查询表内数据。
	DELETE	删除表的数据。
列	SELECT	查询表内的列数据。
函数	ALL	函数的所有操作权限。
	ALTER	修改函数。
	DROP	删除函数。
	DESCRIBE	查看函数的元数据信息。
	EXEC	执行函数。
路径	READ	路径下文件的读权限。
	WRITE	路径下文件的写权限。

隐式 LakeFormation 权限

尽管没有在LakeFormation中进行显式的授权，但是LakeFormation管理员、数据库创建者、表创建者拥有隐式LakeFormation权限。

- LakeFormation管理员：
  - LakeFormation管理员指的是拥有LakeFormation FullAccess权限的用户。
  - 对于其账号下的所有元数据具有读写权限。
  - 可以向任何用户、用户组、角色授予或撤销任何元数据的访问权限。
- 数据库创建者：拥有其创建的数据库的所有数据库权限，拥有其在数据库中创建表的权限，并且可以向同一IAM账号中的其他用户授予在数据库中创建表的权限。数据库创建者对其他人在数据库中创建的表不具有隐式权限。
- 表创建者：
  - 具有其创建的表的所有权限。
  - 可以向同一IAM账号中的主体授予对其创建的所有表的权限。
  - 可以查看包含自己创建的表的数据库。

7 安全

7.1 资产识别与管理

资产识别

- 资产信息：元数据信息、数据权限策略信息。
- 账号信息：不涉及，用户在LakeFormation不感知具体账号信息。
- API映射表：请参见**API参考**。
- LakeFormation云服务需要访问的租户资源包括：租户的用户组和用户信息的读取，对象存储文件目录的创建/删除等，对象存储标签权限接口的访问。

推荐的安全配置

不涉及。

基础设施安全性

- LakeFormation实例运行在跨AZ部署集群，单AZ故障不影响LakeFormation实例的运行。
- LakeFormation实例使用跨AZ高可靠的存储介质来持久化数据，单AZ故障不造成LakeFormation实例的数据丢失。

7.2 身份认证与访问控制

身份认证

- Console界面本租户IAM用户访问LakeFormation。

  LakeFormation针对界面下发的HTTPS请求中IAM Token进行认证，识别出租户、IAM用户等身份。认证失败则拒绝请求。
- Console界面其他租户IAM用户切换到本租户的委托角色来访问LakeFormation。

  LakeFormation针对界面下发的HTTPS请求中IAM Token进行认证，识别出委托方租户、委托、被委托方租户、被委托方IAM用户等身份。认证失败则拒绝请求。

- 其他云服务（如MRS）的实例或集群以本租户的委托身份来访问LakeFormation。

  LakeFormation针对界面下发的HTTPS请求中IAM Token进行认证，识别出委托方租户（本租户）、委托、被委托方租户（ECS云服务账号）、被委托方IAM用户（ECS云服务内置用户）等身份。认证失败则拒绝请求。

资产的访问控制

- 元数据信息的访问控制

  LakeFormation实例针对Console或其他云服务的元数据访问请求，在经过身份认证后，首先进行IAM鉴权，检查用户是否具备请求中的元数据操作权限，接着再进行细粒度的鉴权，检查用户是否具备针对请求中的具体元数据的请求操作权限。鉴权失败则拒绝请求。

- 数据权限策略信息的访问控制

  LakeFormation实例针对Console或其他云服务的元数据访问请求，在经过身份认证后，进行IAM鉴权，检查用户是否具备请求中的权限策略操作权限。鉴权失败则拒绝请求。

7.3 数据保护技术

传输加密（HTTPS）

为保证数据传输的安全性，LakeFormation的API接口为HTTPS协议。

Console或其他云服务需要采用HTTPS协议访问LakeFormation。

数据备份

LakeFormation支持基于其数据备份能力实现LakeFormation实例的数据备份。

7.4 审计

云审计服务（Cloud Trace Service，简称CTS），是华为云安全解决方案中专业的日志审计服务。

CTS可以提供对各种云资源操作记录的收集、存储和查询功能，可用于支撑安全分析、合规审计、资源跟踪和问题定位等常见应用场景。

CTS可用于对LakeFormation实例、元数据等权限的管理。

7.5 更新管理

SSL证书采用SSL协议进行通信，SSL证书部署到服务器后，服务器端的访问将启用HTTPS协议。您的网站将会通过HTTPS加密协议来传输数据，可帮助服务器端和客户端之间建立加密链接，从而保证数据传输的安全。

SSL证书存在有效期限制，证书过期后不被信任，已安装证书的网站业务会受到影响（提示访问不安全或无法访问）。

LakeFormation实例会定期自动更新SSL证书，确保提供持续稳定的HTTPS服务。

8 约束与限制

- 建议使用推荐的浏览器登录LakeFormation管理界面：
  - Chrome：94.0及更高版本。
  - Edge：随Windows操作系统更新。
- IAM用户组被删除后，LakeFormation云服务数据权限中的相关权限策略，需要用户手动清理删除。
- 建库时路径不能为所在Catalog父路径或相同路径，不能为同一Catalog下其他数据库（除default外）路径的父路径、子路径或相同路径。
- 创建数据库的存储位置必须在所属Catalog的存储位置之下。
- 用户自定义创建的Catalog对象及其子元数据对象，暂不支持授权和细粒度权限控制。
- LakeFormation数据权限单次授权，授权主体不能超过20个，元数据对象不能超过10个。
- LakeFormation中，总分区数量不超过1,000,000,000。
- LakeFormation不支持跨Region的元数据和权限统一管理。
- LakeFormation不支持跨实例的元数据和权限统一管理。
- 一个数据表中，每个分区所对应的Partition Value组合必须在全表唯一。
- 由Partition Keys和Partition Values组合构成的Partition Name，总长度不能超过1000字符。
- 元数据的参数描述中，1个中文字符对应3个字节。
- LakeFormation依赖OBS服务的并行文件系统，OBS需要基于大数据存算分离场景进行分离部署；LakeFormation元数据的存储位置对应OBS路径，与存算分离架构的MRS等大数据集群对接。OBS并行文件系统需要支持AccessLabel特性。
- LakeFormation中，不同实例的同名角色在授权时对应的OBS AccessLabel相同，不建议在同一个区域中的不同实例创建同名角色。

9 产品生命周期

生命周期是指LakeFormation实例从创建到删除（或释放）历经的各种状态。LakeFormation各状态说明请参考**表9-1**。

表 9-1 LakeFormation 状态说明

状态	说明
资源准备中	创建LakeFormation实例后，LakeFormation实例状正在进行资源准备。
资源准备失败	创建LakeFormation实例后，LakeFormation实例准备资源失败。
运行中	LakeFormation实例正常运行状态。在这个状态的实例可以运行您的业务。
资源释放中	执行删除LakeFormation实例操作后，资源正在进行释放。
删除中	触发删除LakeFormation实例后，在LakeFormation实例在彻底被删除之前的状态。
已删除	LakeFormation实例已经删除成功。
恢复中	已删除的实例正在从回收站中恢复。
冻结	如果您的账号因为欠费或者违规，LakeFormation实例将被冻结。此时该实例处于只读状态，不能进行修改和删除操作。

10 与其他服务的关系

LakeFormation服务与其他服务的关系如下表所示。

表 10-1 LakeFormation 服务与其他服务的关系

服务名称	LakeFormation服务与其他服务的关系
统一身份认证（Identity and Access Management，IAM）	通过IAM完成对IAM用户或委托的身份认证以及部分访问控制。
云审计服务（Cloud Trace Service，CTS）	云审计服务记录LakeFormation服务相关的操作事件，方便用户日后的查询、审计和回溯。
对象存储服务（Object Storage Service，OBS）	LakeFormation服务的元数据所映射的实际业务数据，存储在OBS并行文件系统的目录和文件。
MapReduce服务（MapReduce Service，MRS）	LakeFormation与MRS集群中的Ranger、Hive、Spark对接，实现湖、仓元数据统一管理。
数据仓库服务 GaussDB（DWS）	LakeFormation与DWS对接，实现湖、仓元数据统一管理。