

数据工坊(DWR)

产品介绍

文档版本 03
发布日期 2023-05-06



版权所有 © 华为技术有限公司 2023。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。

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

<https://www.huawei.com/cn/psirt/vul-response-process>

如企业客户须获取漏洞信息，请参见如下网址：

<https://securitybulletin.huawei.com/enterprise/cn/security-advisory>

目录

1 什么是数据工坊.....	1
2 产品优势.....	3
3 应用场景.....	4
4 使用限制.....	5
5 计费说明.....	6
6 权限管理.....	7
7 与其他服务的关系.....	11
8 基本概念.....	13
9 修订记录.....	14

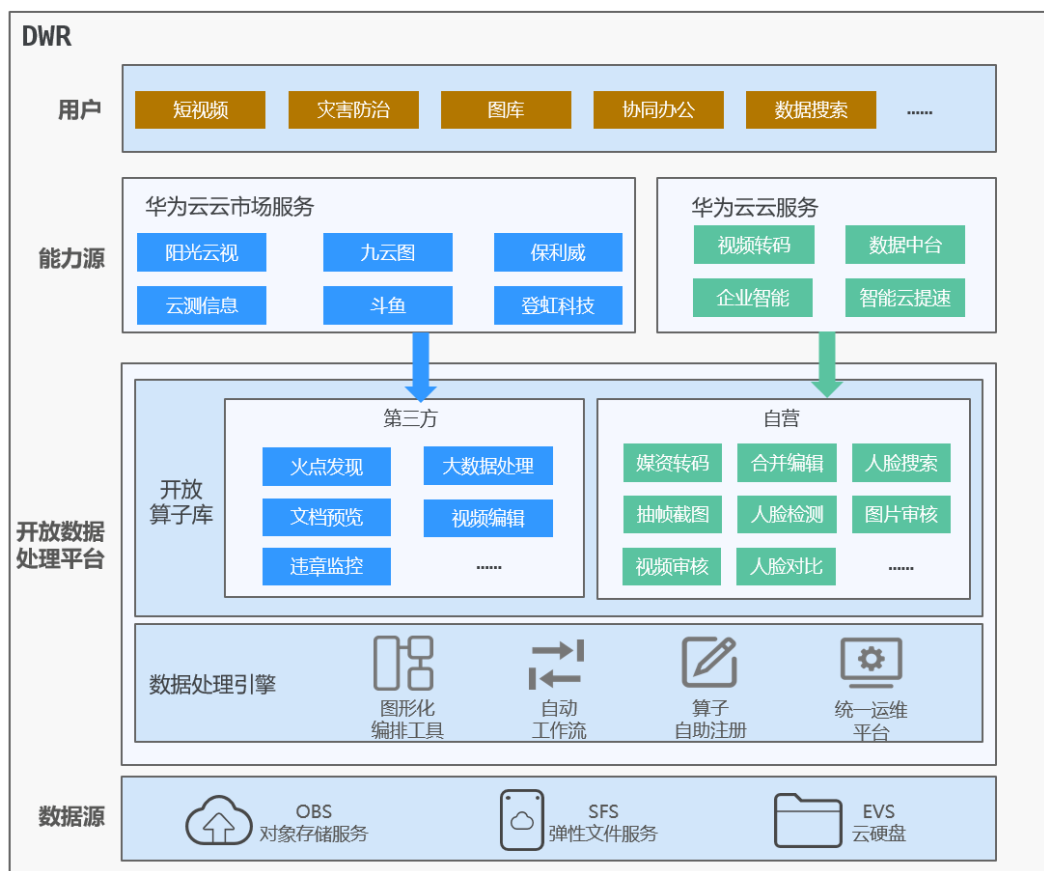
1 什么是数据工坊

数据工坊（Data Workroom，DWR）是一款近数据处理服务，通过易用的数据处理 workflow 编排和开放生态的数据处理算子，能够在云上实现图像、视频、文档、图片等数据处理业务。

如图1-1所示，DWR主要对外提供开放的算子库和数据处理引擎。

- 算子库包含由DWR提供的华为自有算子和第三方开发者提供的第三方算子。
 - 华为自有算子的能力源是华为云数据处理相关的云服务，如媒体处理MPC、图像识别Image等，DWR将云服务提供的各种数据处理能力通过函数生成算子集成在算子库中。
 - 第三方算子是基于DWR的算子注册能力，由第三方开发者创建，专业人员审核发布的公共算子，您也可以将自己创建的算子发布为第三方算子，开放给所有华为云用户使用。
- 数据处理引擎的核心是DWR提供的工作流图形化编排能力，用户可以通过图形化的界面，将任意算子编排到工作流中。通过事件触发器或API驱动，DWR将根据您定义的工作流自动进行数据处理。

图 1-1 DWR 架构示意图



2 产品优势

数据处理方式对比

- 传统线下处理方式：硬件为用户自建IDC，软件为自研或集成商的数据处理软件，通过数据处理软件完成数据处理。
- 传统云上处理方式：使用云上存储服务和数据处理服务，数据写入存储服务后，再调用数据处理服务接口实现数据处理。
- 云上近数据处理方式：使用云上存储服务和DWR的近数据处理能力，数据写入存储服务后，自动触发DWR编排的工作流完成数据处理。

表 2-1 数据处理方式对比

对比项	传统线下处理	传统云上处理	云上近数据处理
成本	软硬件自建自维，成本高	资源按需购买，自建或使用云上服务，成本较低	资源按需购买，开放算子库，成本低
效率	独立组件无优化，效率低	端到端垂直优化，数据处理路径长，效率较高	端到端垂直优化，算子下推路径短，效率高
复杂度	业务流程复杂，复杂度高	业务流程复杂，复杂度高	图形化编排工具，一站式运维平台，复杂度低

DWR 的优势

- **高效：数据处理时间缩短60%**
近数据计算：调用算子提供商设定好的云服务器API来执行算子，函数化调用，路径短。
- **易用：开发运维效率提升50%+**
图形化编排：拖拉式智能模块化操作，免代码。
一站式运维：可视化统一运维平台，降低管理难度。
- **开放：算子库能力无上限**
丰富算子生态：可调用海量华为云自有算子及第三方算子，免开发。
开放式自助注册：第三方算子轻松上架，引流增收。

3 应用场景

未来几年时间，数据量将会呈数倍增长趋势，海量数据的快速处理是企业数据变现的核心诉求。DWR提供的高效、开放、易用的近数据处理能力，让视频处理、图片审核、文档压缩、灾难防治、数据搜索和推荐、办公协同、基因测序等场景的海量数据快速产生价值。

DWR的应用场景包括但不限于以下典型场景。

表 3-1 DWR 典型应用场景

应用场景	场景下支持的能力
视频	视频解析、媒资转码、抽帧截图、添加水印、转封装、审核等
图片	缩略、添加水印、转封装、文字审核、图像审核、人脸识别等
文档	解压、预览等
基因	文件个数统计、模糊搜索、压缩、容量监控通知等
医疗	数据从存储网关到OBS、AI分析、影像筛查等
监控	视频截帧、图片压缩、图片转存、车牌识别、人脸识别、RTMP直存OBS等

4 使用限制

区域限制

当前支持使用的区域为：北京四、上海一、上海二、广州、新加坡。

使用限制

- 当前仅支持OBS对象桶，不支持并行文件系统。
- 当前仅支持创建串行的工作流。
- 一个事件触发器仅支持关联一个工作流，一个工作流可同时被多个事件触发器关联。
- 当前暂不支持创建重名工作流，即使工作流被删除后，也不能再创建与之重名的工作流。
- 同一桶内的触发器名称不允许重复。
- 工作流及工作流中使用的FunctionGraph、数据处理服务、消息通知服务、OBS桶等均需要在相同区域的默认项目下。
- 细粒度授权不支持企业项目。
- 一个桶支持绑定10个事件触发器。
- 当前发布算子功能仅支持IAM主账号。

5 计费说明

算子费用主要由三部分组成：OBS API调用费、FunctionGraph函数及函数 workflow 费、算子费。

表 5-1 计费项

计费项	说明
OBS API调用费	算子请求OBS API的调用费。 通过算子对数据进行处理，都会涉及到对OBS API的调用，每调用一次API都计算一次请求次数。对象存储服务OBS会根据调用API的请求次数进行费用收取，收取详情参见 OBS请求费用说明 。
FunctionGraph函数及函数 workflow 费	算子使用FunctionGraph函数 workflow 的资源费用。 通过算子对数据进行处理，会使用到函数 workflow 的资源，比如算子执行时长，函数 workflow 会根据资源使用情况进行收费，收费详情参见 函数 workflow 计费说明 。
算子费	算子本身的费用。 <ul style="list-style-type: none">第三方算子在发布之前需要先在云市场上架，云市场上算子的收费标准由算子发布者决定。自营算子由算子云服务提供方决定收费标准。例如使用人脸识别算子，该算子由人脸识别服务FRS提供，会根据人脸识别接口调用次数进行收费。查询算子的云服务提供方参见表华为算子一览，各个云服务的计费规则请查看该服务的“计费说明”章节。

6 权限管理

如果您需要对华为云上购买的DWR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，并授权控制他们对华为云资源的访问范围。例如您的员工中有负责软件开发的人员，您希望他们拥有DWR的使用权限，但是不希望他们拥有删除DWR等高危操作的权限，那么您可以使用IAM为开发人员创建用户，通过授予仅能使用DWR，但是不允许删除DWR的权限，控制他们对DWR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，您可以跳过本章节，不影响您使用DWR服务的其它功能。

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](#)。

DWR 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。

DWR部署时通过物理区域划分，为项目级服务。授权时，“作用范围”需要选择“区域级项目”，然后在指定区域（如华北-北京1）对应的项目（cn-north-1）中设置相关权限，并且该权限仅对此项目生效；如果在“所有项目”中设置权限，则该权限在所有区域项目中都生效。访问DWR时，需要先切换至授权区域。

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DWR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。多数细粒度策略以API接口为粒度进行权限拆分。

如表6-1所示，包括了DWR的所有系统权限。

表 6-1 DWR 系统权限

系统角色/策略名称	描述	类别
Tenant Administrator	拥有该权限的用户拥有除IAM外，其他所有服务的所有执行权限。	系统角色
DWR FullAccess	管理员权限，拥有该权限的用户可以操作并使用所有操作。	系统策略
DWR ReadOnlyAccess	只读权限，拥有该权限的用户仅能查看DWR服务数据。	系统策略

DWR操作与资源权限关系列出了DWR常用操作与系统权限的授权关系，您可以参照该表选择合适的系统权限。

表 6-2 DWR 操作与资源权限关系

操作	DWR FullAccess	DWR ReadOnlyAccess
查询 workflow 实例详情	√	√
查询 workflow 详情	√	√
查询华为云系统算子	√	√
查询公共算子	√	√
查询用户第三方算子	√	√
查询第三方算子服务协议是否同意	√	√
获取DWR workflow 是否开通授权	√	√
更新 workflow 参数	√	x
更新用户第三方算子	√	x
恢复的 workflow 实例	√	x
DWR workflow 开通授权	√	x
同意第三方算子服务协议	√	x
列举 workflow	√	√
列举华为云系统算子	√	√
列举公共算子	√	√
列举用户第三方算子	√	√

操作	DWR FullAccess	DWR ReadOnlyAccess
禁用用户第三方算子	√	x
创建工作流	√	x
删除工作流	√	x
删除用户第三方算子	√	x
创建用户第三方算子	√	x
执行工作流	√	x

授权

使用DWR之前需要进行相关授权，才能保证用户正常使用DWR，并且与其他数据处理服务相互配合进行数据处理。主要分以下两个方面：

1. 第一次使用DWR时，需要在控制台上根据提示对DWR进行统一授权。

授权内容：

授权DWR使用您的FunctionGraph服务和DWR工作流中使用到的其他数据处理服务，用以创建函数和通过函数调用其他数据处理服务的API，以便能根据工作流自动处理OBS桶中的数据。

授权过程用户无感知，只需一键操作，系统将自动为您创建IAM委托。

- 委托一：委托OBS服务访问FunctionGraph服务
- 委托二：委托FunctionGraph服务访问OBS服务以及其他数据处理服务

📖 说明

- 在授权前，需要保证账号下至少还可以创建2个委托。
 - 访问其他数据处理服务的权限不会在统一授权时添加，系统会在您创建工作流时，根据工作流函数使用到的服务按需添加。
 - 如果账号的下的IAM用户需要使用DWR，则需要给IAM用户配置FunctionGraph、数据处理服务（如MPC）、SMN等服务的访问权限。具体配置的权限需要和委托的权限保持一致。委托的权限详情可以在“统一身份认证 > 委托”中查看名称以“obs_function_agency”开头的委托。
2. 使用媒体处理MPC的相关能力时（如视频解析、抽帧截图、媒资转码等），需要先在MPC中开启桶授权。授权方法请参见[获取云资源授权](#)。

需要为以下桶开启桶授权：

- 抽帧截图输出桶
- 媒资转码输出桶
- 使用了视频解析、抽帧截图或媒资转码的工作流所关联的事件触发器对应的事件源存储桶

📖 说明

您也可以直接开启委托授权，一键为所有桶都进行授权。但为了遵循权限最小原则，建议您选择通过桶授权方式仅为有需求的桶授权。

相关链接

- [IAM产品介绍。](#)
- [创建用户组、用户并授予DWR权限。](#)

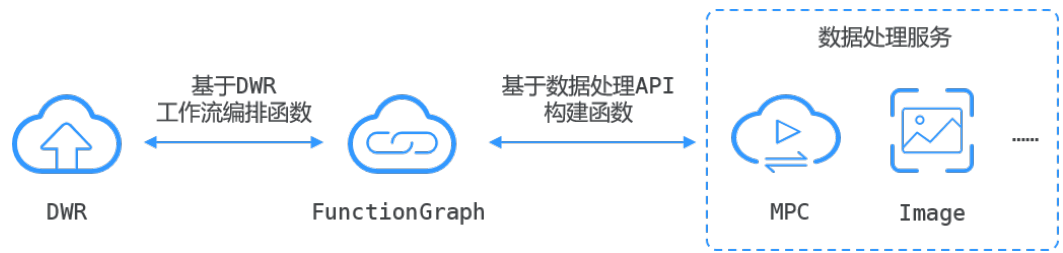
7 与其他服务的关系

DWR数据处理的数据源是华为云存储服务，DWR提供的华为云自有算子是通过函数生成，且能力源是华为云数据处理相关的云服务。因此，DWR与其他服务的关系如表7-1所示。

表 7-1 DWR 与其他云服务的关系

服务类别	服务名称	交互关系
存储服务	对象存储服务 (Object Storage Service, OBS)	OBS作为DWR数据处理的数据源，当上传至OBS或存储在OBS的数据产生的事件满足触发器规则时，DWR将自动启动关联的工作流进行数据处理。
函数服务	函数工作流 (FunctionGraph)	DWR中由华为云提供的自有算子均是利用数据处理服务的API，通过FunctionGraph封装而成。
数据处理服务	媒体处理 (Media Processing Center, MPC)	DWR提供的视频解析、抽帧截图、媒资转码等算子，依赖MPC提供的相应能力。
	图像识别 (Image Recognition)	DWR提供的图像标签算子，依赖Image Recognition提供的相应能力。
	视频分析服务 (Video Analysis Service, VAS)	DWR提供的视频内容审核算子，依赖VAS提供的相应能力。
	DWR会根据华为云数据处理服务的能力，不断提供新的算子，此处不再穷举。

图 7-1 DWR 与函数 workflow 及数据处理服务的关系



8 基本概念

workflow

workflow是DWR运行单位，其中可以包含一个或多个算子，客户可通过加入不同算子到 workflow中实现高度自由化的数据处理功能。

算子

算子是DWR中的执行单位，是一个具有输入输出的完成功能模块。算子既有华为云提供的内置官方算子，也有经过华为云审核后的第三方算子。用户可以根据自己的需求挑选算子加入 workflow中，从而快速构建特有的数据处理功能。

触发器

每个 workflow需要绑定一个触发器用于触发执行 workflow，该触发器当前仅支持作用于 OBS桶上，用户可在触发器中定义 workflow的触发方式。

异步启动 workflow

调用用户的工作流，并且客户端不关注请求调用的结果，服务端收到请求后将请求排队，排队成功后请求就返回，服务端在空闲的情况下会逐个处理排队的请求。

同步启动 workflow

调用用户的工作流，并且客户端请求需要明确等到响应结果，也就是说同步调用需等待后台处理完成后返回结果。不同于异步的请求，同步请求在等待期间请求将一直阻塞。

9 修订记录

发布日期	修订记录
2023-05-06	第三次正式发布。 本次变更说明如下： 使用限制 ，新增不支持并行文件系统的限制。
2022-04-30	第二次正式发布。 本次变更说明如下： 增加IAM细粒度授权。
2022-02-14	第一次正式发布。