

阿里云 DataWorks

产品简介

文档版本：20190828

法律声明

阿里云提醒您在使用或阅读本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 禁止： 重置操作将丢失用户配置数据。
	该类警示信息可能导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告： 重启操作将导致业务中断，恢复业务所需时间约10分钟。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明： 您也可以通过按Ctrl + A选中全部文件。
>	多级菜单递进。	设置 > 网络 > 设置网络类型
粗体	表示按键、菜单、页面名称等UI元素。	单击 确定 。
<code>courier</code> 字体	命令。	执行 <code>cd /d C:/windows</code> 命令，进入Windows系统文件夹。
<code>##</code>	表示参数、变量。	<code>bae log list --instanceid</code> <code>Instance_ID</code>
<code>[]</code> 或者 <code>[a b]</code>	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
<code>{ }</code> 或者 <code>{a b}</code>	表示必选项，至多选择一个。	<code>swich {stand slave}</code>

目录

法律声明.....	I
通用约定.....	I
1 什么是DataWorks.....	1
2 基本概念.....	3
3 应用场景.....	7
4 数据开发流程.....	10
5 简单模式和标准模式的区别.....	11
6 版本历史.....	14
7 文档最新动态.....	21

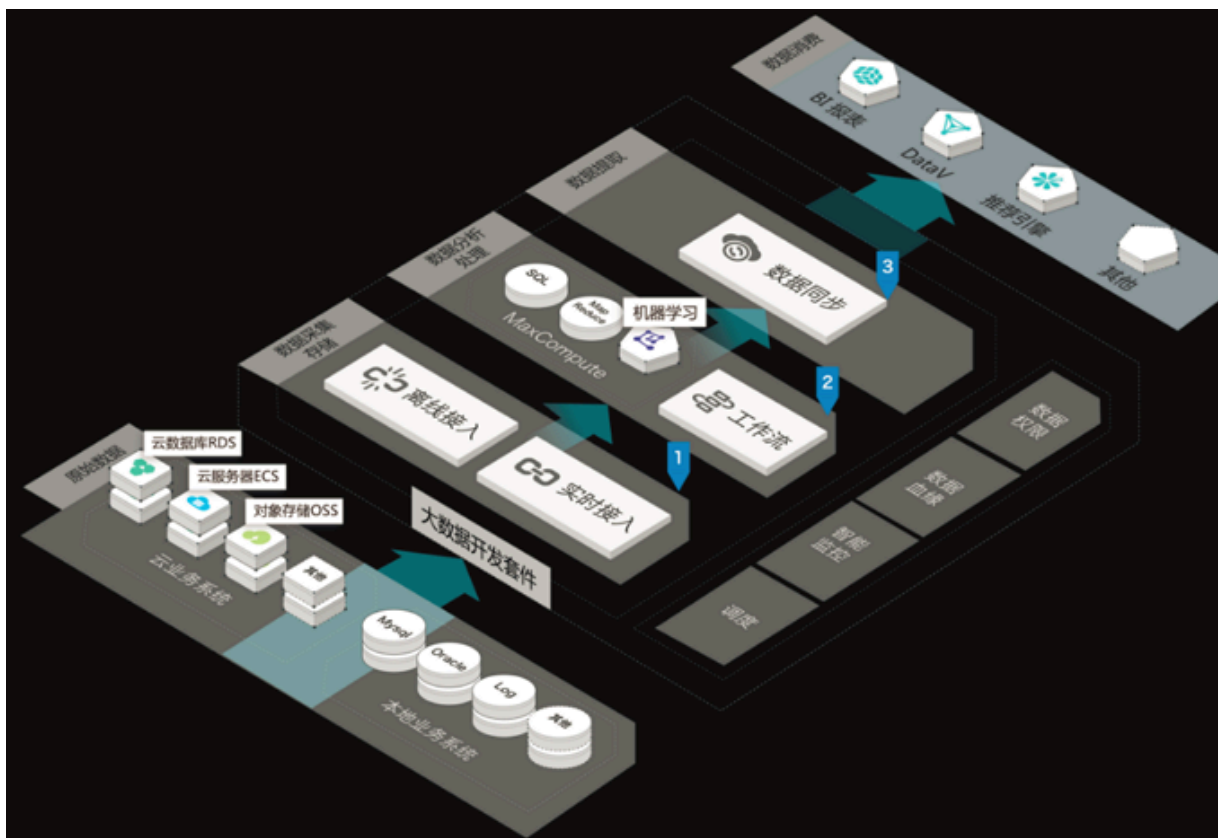
1 什么是DataWorks

本文将为您介绍什么是DataWorks，以及DataWorks的功能、约束与限制。

DataWorks（数据工场，原大数据开发套件）是阿里云重要的PaaS平台产品，为您提供**数据集成**、**数据开发**、**数据管理**、**数据质量**和**数据服务**等全方位的产品服务，一站式开发管理的界面，帮助企业专注于数据价值的挖掘和探索。

DataWorks基于MaxCompute作为核心的计算、存储引擎，为您提供海量数据的离线加工分析、数据挖掘等功能，详情请参见[MaxCompute简介](#)。

您可以通过DataWorks，对数据进行传输、转换和集成等操作。从不同的数据存储引入数据，并进行转化和开发，最后将处理好的数据同步至其他数据系统。



学习路径

您可以通过[DataWorks学习路径](#)，快速了解DataWorks V2.0的相关概念、基础操作和进阶操作等内容。

功能概述

- 全面托管的调度
 - DataWorks提供强大的调度能力，详情请参见[调度配置](#)。
 - 支持根据时间、依赖关系，进行任务触发的机制。详情请参见[时间属性](#)和[依赖关系](#)。
 - 支持每日千万级别的任务，根据DAG关系准确、准时地运行。
 - 支持分钟、小时、天、周和月多种调度周期配置。
 - 完全托管的服务，无需关心调度的服务器资源问题。
 - 提供隔离功能，确保不同租户之间的任务不会相互影响。
- 支持多种节点类型，详情请参见[节点类型](#)。

DataWorks支持[数据同步](#)、[Shell](#)、[MaxCompute SQL](#)、[MaxCompute MR](#)等多种节点类型，通过节点之间的相互依赖，对复杂的数据进行分析处理。

- [数据转化](#)：依托MaxCompute强大的能力，保证了大数据的分析处理性能。
 - [数据同步](#)：依托DataWorks中[数据集成](#)的强力支撑，支持超过20种[数据源](#)，为您提供稳定高效的数据传输功能。
- 可视化开发

DataWorks提供[可视化](#)的代码开发、工作流设计器页面，无需搭配任何开发工具，简单拖拽和开发，即可完成复杂的数据分析任务。只要有浏览器有网络，便可随时随地进行开发工作。

- 监控告警

运维中心提供可视化的任务监控管理工具，支持以DAG图的形式展示任务运行时的全局情况，详情请参见[运维中心](#)。

您可以方便地配置各类报警方式，任务发生错误可及时通知相关人员，保证业务正常运行。详情请参见[智能监控](#)。

约束与限制

- 仅支持Chrome浏览器54以上版本。
- 目前无法支持SQL运行在阿里云云数据库、分析型数据库等产品，仅支持MaxCompute。

2 基本概念

本文将为您介绍DataWorks中，业务流程、解决方案、组件、任务、实例、提交、脚本开发、资源、函数和输出名称等基本概念。

业务流程

针对业务实体，抽象出业务流程的概念，帮助您从业务视角组织代码的开发，提高任务管理效率。



说明：

业务流程可以被多个解决方案复用。

业务流程的优势：

- 帮助您从业务视角组织代码。
- 支持基于任务类型的代码组织方式。
- 支持多级子目录（建议不超过四级）。
- 支持从业务视角查看整体的工作流，并进行优化。
- 提供业务流程看板，帮助您更高效地进行开发。
- 支持根据业务流程组织发布和运维。

解决方案

您可以自定义组合部分业务流程为一个解决方案。

解决方案的优势：

- 一个解决方案包括多个业务流程。
- 解决方案之间可以复用相同的业务流程。
- 组织完成的解决方案，可以进行沉浸式开发。

组件

组件是带有多个输入参数和输出参数的SQL代码过程模板。SQL代码的处理过程通常是引入一到多个源数据表，通过过滤、连接和聚合等操作，加工出新的业务需要的目标表。

您可以将SQL中的通用逻辑抽象为组件，提高代码的复用性。

任务 (Task)

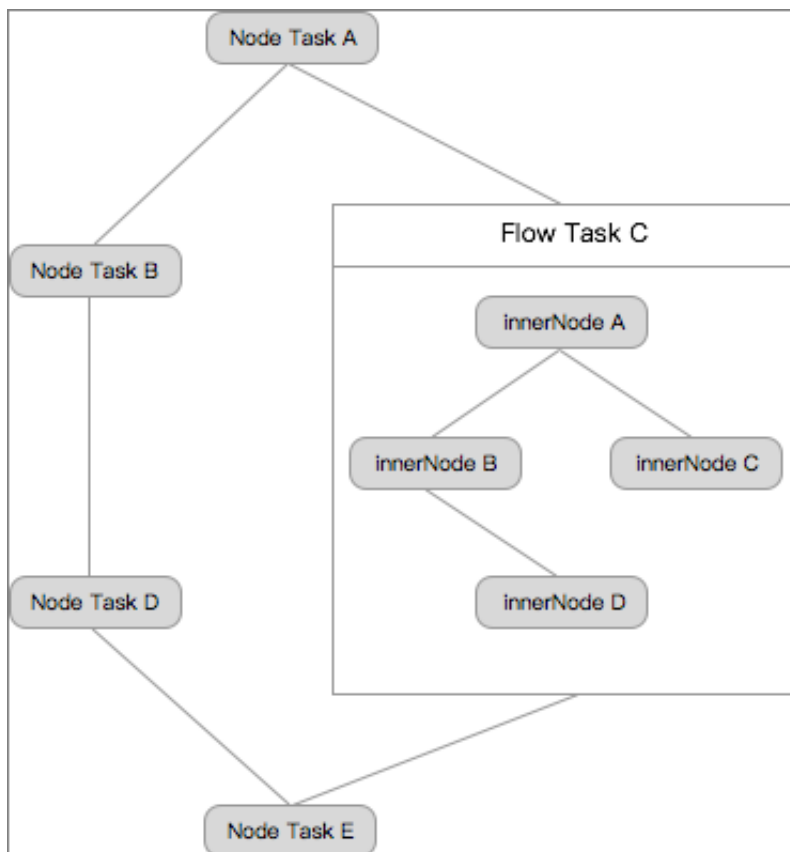
任务是对数据执行的操作的定义。示例如下：


- 通过数据同步节点任务，将数据从RDS同步至MaxCompute。
- 通过MaxCompute SQL节点任务，运行MaxCompute SQL来进行数据的转换。

- 通过 workflow 任务中不同的 SQL 内部节点，完成一系列的数据转换。

每个任务使用 0 或 0 个以上的数据表（数据集）作为输入，生成一个或多个数据表（数据集）作为输出。

任务主要分为节点任务（node task）、workflow 任务（flow task）和内部节点（inner node）。



任务类型	说明
节点任务（node task）	一个数据执行的操作。可以与其他节点任务、workflow 任务配置依赖关系，组成 DAG 图。
workflow 任务（flow task）	<p>满足一个业务场景需求的一组内部节点，组成一个 workflow 任务，建议 workflow 任务小于 10 个。</p> <p>workflow 任务内部节点，无法被其他 workflow 任务、节点任务依赖。workflow 任务可以与其他 workflow 任务、节点任务配置依赖关系，组成 DAG 图。</p> <p> 说明： 从 DataWorks V1.0 升级的任务，仍保留 workflow 的概念。DataWorks V2.0 已无法创建 workflow 任务，您可以选择创建业务流程进行后续操作。</p>
内部节点（inner node）	workflow 任务内部的节点，与节点任务的功能基本一致。其调度周期会继承 workflow 任务的调度周期，无法进行单独配置，通过拖拽形成依赖关系。

数据执行时可以选择的操作类型，请参见[节点类型](#)。

任务的调度参数配置，请参见[调度配置](#)。

实例 (Instance)

实例是某个任务在某时某刻执行的一个快照。调度系统中的任务，经过调度系统、手动触发运行后，会生成一个实例。实例中会有任务的运行时间、运行状态和运行日志等信息。

例如设置每天2:00运行task1实例，调度系统会在每天23:30根据周期节点定义好的时间，自动生成一个快照，即task1第二天2:00运行的实例。到第二天2:00时，如果判断上游实例已经完成，task1实例便会如期启动运行。



说明:

您可以进入[运维中心 > 任务运维页面](#)，查询实例的相关信息。

提交 (Submit)

提交是指开发的节点任务、工作流任务，从DataWorks开发环境发布至调度系统的过程。完成提交后，相应的代码、调度配置全部合并至调度系统中，调度系统根据相关配置进行调度操作。



说明:

未提交的节点任务、工作流任务不会进入调度系统。

脚本开发 (Script)

脚本开发是提供给数据分析使用的一个代码存储空间。脚本开发的代码无法发布到调度系统，无法进行调度参数配置，仅可以进行部分数据查询分析的工作。

资源、函数

资源、函数均为MaxCompute的概念，详情请参见[#unique_23](#)和[#unique_24](#)。

您可以在DataWorks中，通过界面管理资源和函数。如果通过MaxCompute的其他方式进行资源、函数管理，则无法在DataWorks中进行相关的查询。

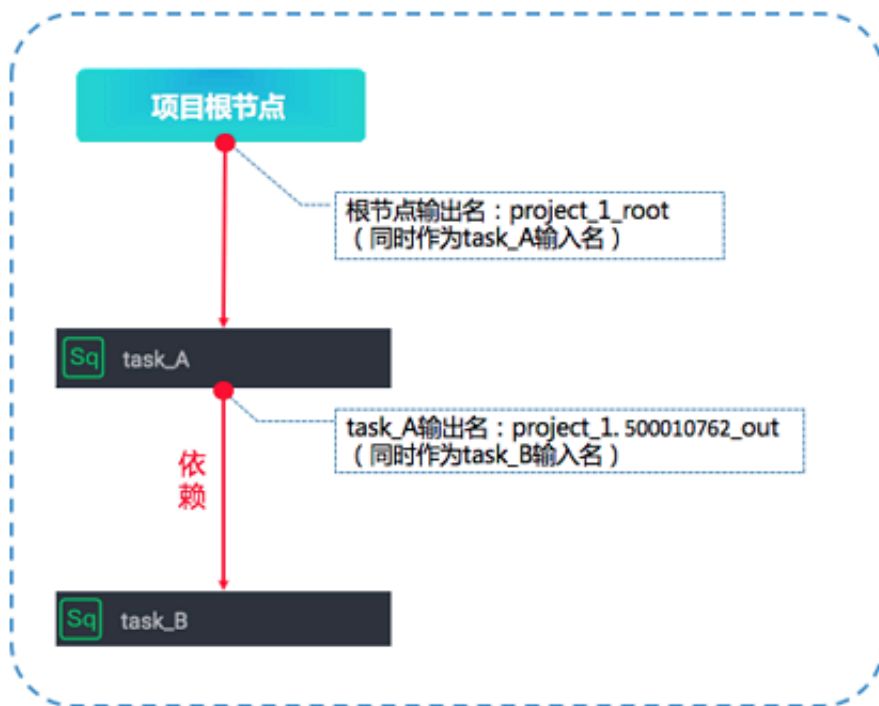
输出名称

输出名称：每个任务 (Task) 输出点的名称。它是您在单个租户 (阿里云帐号) 内设置依赖关系时，用于连接上下游两个任务 (Task) 的虚拟实体。

当您在设置某任务与其他任务形成上下游依赖关系时，必须根据输出名称 (而不是节点名称或节点ID) 来完成设置，设置完成后该任务的输出名也同时作为其下游节点的输入名称。



Project_1



说明:

输出名称可以作为某个Task在同租户内，区别于其他Task的唯一概念对象，每个节点的输出名称默认为项目名称.系统生成9位数字_out。您可以对Task增加自定义输出名，但需注意输出节点名称在租户内不允许重复。

3 应用场景

本文将为您介绍DataWorks的应用场景示例。

日志大数据分析

- 提高工作效率。

将日志数据同步至MaxCompute，通过SQL语句进行分析与处理，提高工作效率。

- 提高存储利用率。

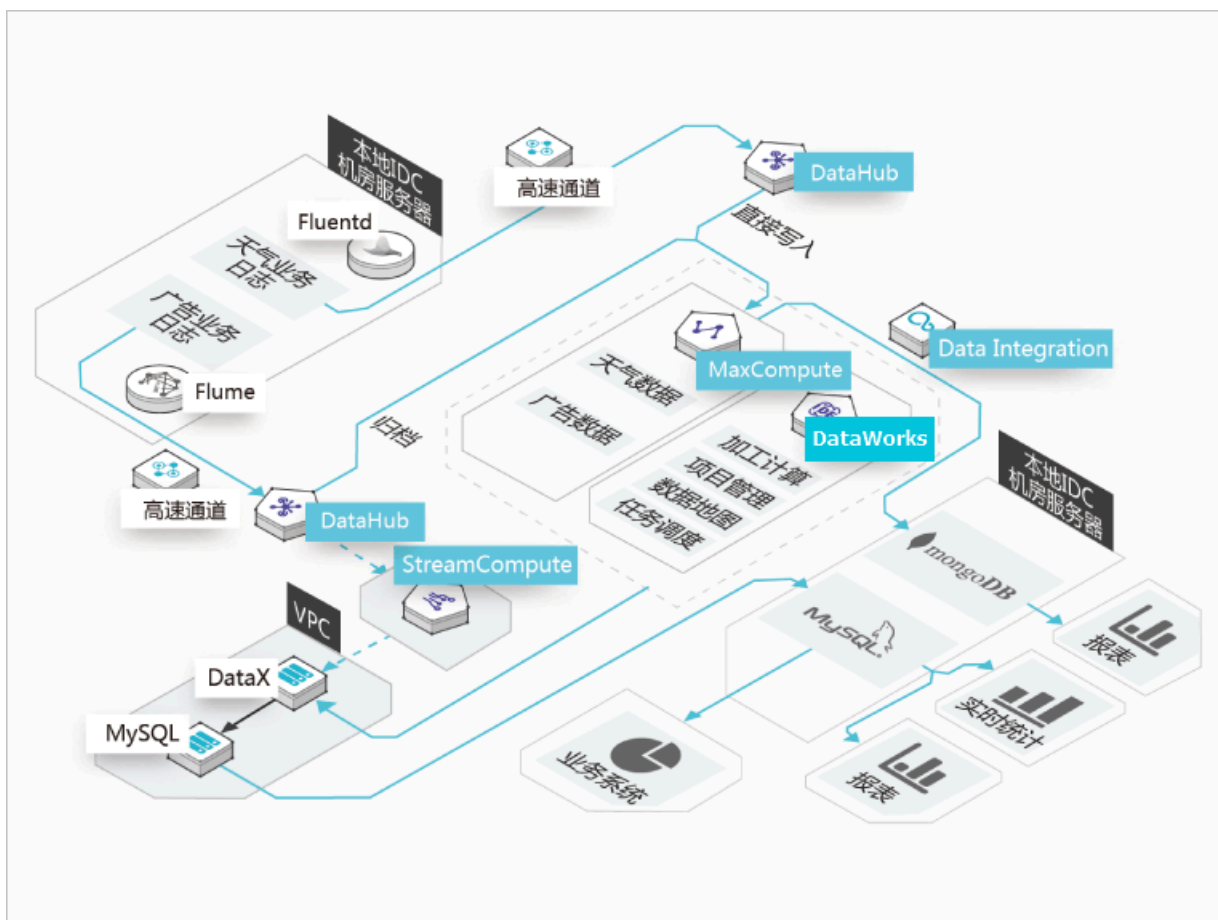
可以降低整体存储和计算的费用，并提高性能和稳定性。

- 降低大数据使用门槛。

MaxCompute提供多种开源软件的插件，可以轻松完成数据上云。

推荐搭配使用：

DataWorks + 数据集成 + AnalyticDB for MySQL + Quick BI + MaxCompute



客户案例请参见[墨迹天气](#)。

精细化运营

- 提升业务洞察能力。

通过MaxCompute计算能力，可以实现针对百万用户的精细化运营。

- 业务数据化。

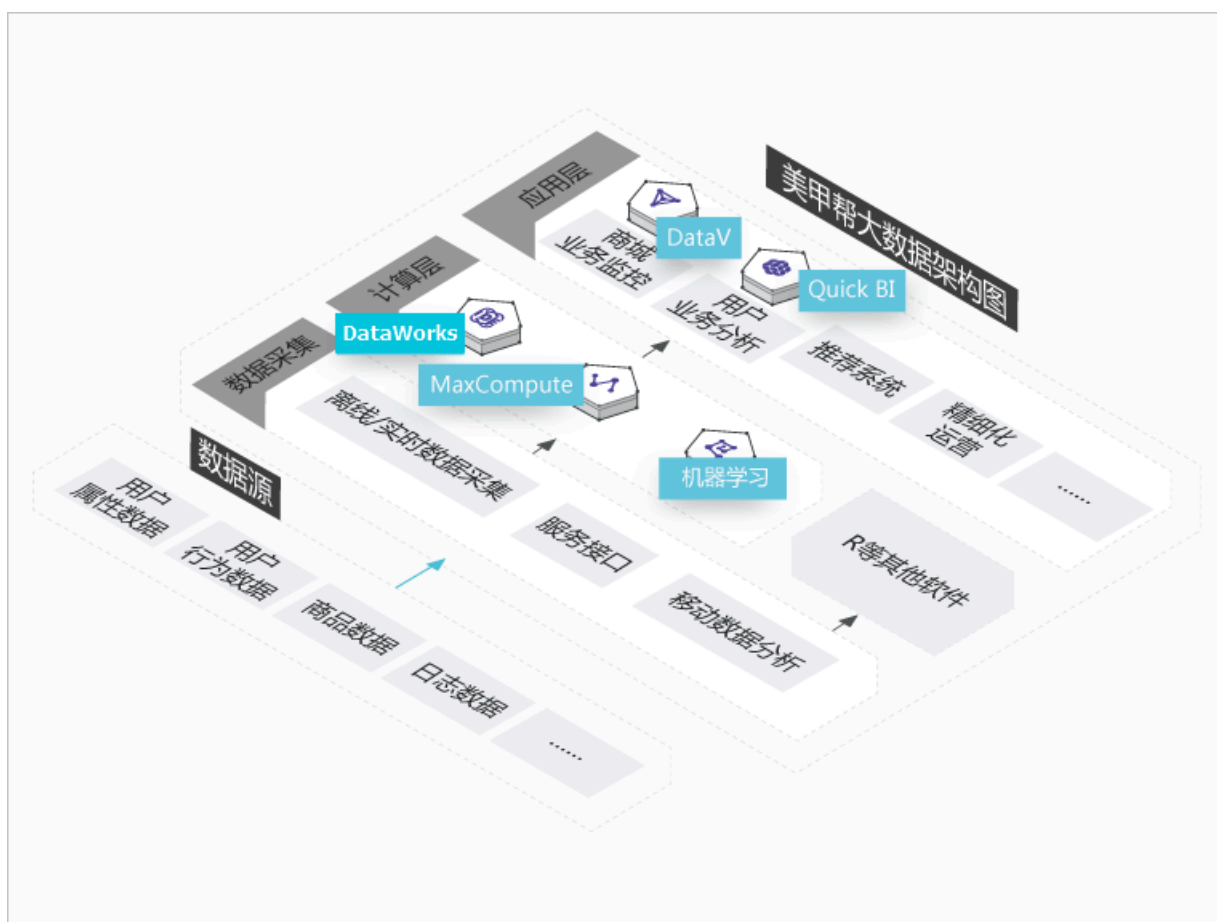
可以提升对业务数据的分析能力并进行有效监控，更好地业务赋能。

- 快速响应业务需求。

可以根据新业务的数据分析需求，快速灵活地进行响应与满足。

推荐搭配使用：

DataWorks + 数据集成 + Quick BI + MaxCompute



客户案例请参见[美甲帮](#)。

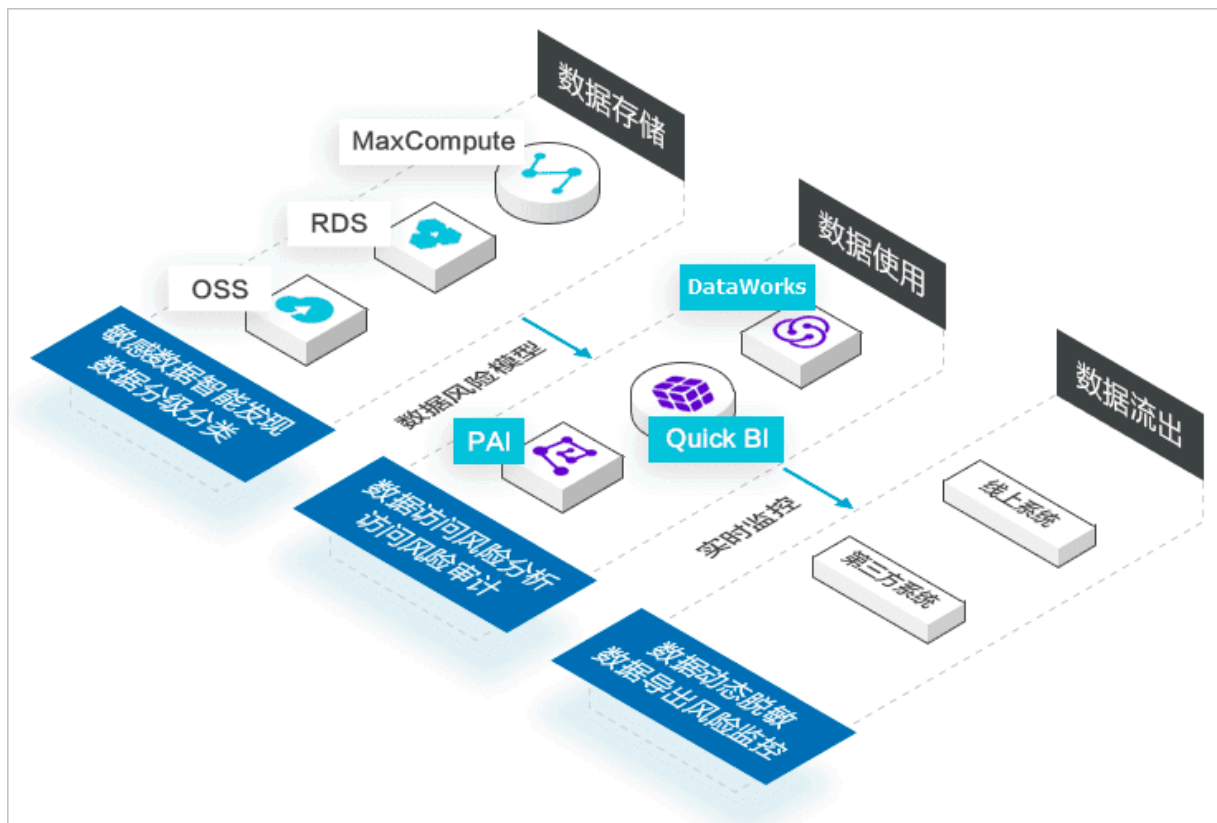
数据安全治理

- 敏感数据识别。

通过用户自定义规则，自动识别敏感数据，并标记对应的级别。

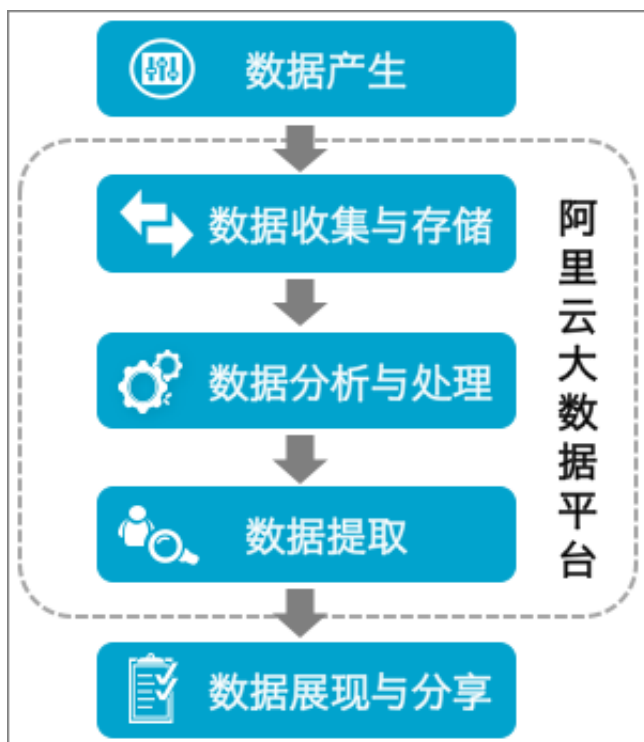
- 敏感数据展示脱敏。
提供设置脱敏规则功能，实现敏感数据查询展示脱敏。
- 敏感数据操作风险监控。
可视化监控数据分布、数据使用和数据导出，提供自定义风险识别和审计功能。

推荐使用DataWorks的数据保护伞功能。



4 数据开发流程

通常情况下，数据开发的总体流程包括数据产生、数据收集与存储、数据分析与处理、数据提取和数据展现与分享。



说明：

上图中，虚线框内的开发流程均可基于阿里云大数据平台完成。

1. **数据产生**：业务系统每天会产生大量结构化的数据，存储在业务系统所对应的数据库中，包括MySQL、Oracle和RDS等类型。
2. **数据收集与存储**：如果想通过MaxCompute的海量数据存储与处理能力分析已有的数据，需要首先将不同业务系统的数据同步至MaxCompute中。

DataWorks提供数据集成服务，可以支持多种数据源类型，将业务系统数据根据预设的调度周期同步至MaxCompute。

3. **数据分析与处理**：可以对MaxCompute上的数据进行加工（MaxCompute SQL、MaxCompute MR）、分析与挖掘（数据分析、数据挖掘）等处理，从而发现其价值。
4. **数据提取**：分析与处理后的结果数据，需同步导出至业务系统，以供业务人员使用其分析的价值。
5. **数据展现与分享**：数据提取成功后，可以通过报表、地理信息系统等多种展现方式，展示与分享大数据分析、处理后的成果。

5 简单模式和标准模式的区别

Dataworks V2.0为您提供简单模式和标准模式两种工作空间模式，本文将为您介绍简单模式和标准模式的区别。

简单模式

简单模式指一个Dataworks工作空间对应一个MaxCompute项目，无法设置开发环境和生产环境，只能进行简单的数据开发。简单模式的工作空间无法对数据开发流程和表权限进行强控制。

简单模式的优势：迭代快，代码提交后，无需发布即可生效。

简单模式的风险：开发角色权限过大，可以随意删除本工作空间下的表，导致表权限存在风险。

标准模式

标准模式指一个Dataworks工作空间对应两个MaxCompute项目，您可以设置开发和生产双环境，提升代码开发规范。

标准模式的工作空间可以严格控制表权限，禁止随意操作生产环境的表，保证生产表的数据安全。

创建工作空间

基本信息

workspace名称:

显示名:

* 模式: 标准模式 (开发跟生产隔离) ▼

描述:

高级设置

* 启动调度周期: 开 ?

* 能下载select结果: 开 ?

面向 MaxCompute

* MaxCompute项目名称: ?

* MaxCompute访问身份: 工作空间所有者 ?

* Quota组切换: ▼

- 所有任务只能在开发环境进行编辑，无法直接修改生产环境的代码，保证生产环境代码的稳定。
- 开发环境默认不开启任务调度，避免开发环境的任务周期运行，和生产环境的任务抢夺资源，更好地保障生产环境任务运行的稳定性。
- 生产环境运行会有一个默认的生产账号，生产账号产出的所有表都属于主账号。如果开发过程中需要使用生产表，必须单独申请权限，更好地控制表权限。

创建工作空间时，选择模式为标准模式（开发跟生产隔离），填写工作空间名称和描述，其余配置项选择默认值即可。



说明:

- 生产环境的MaxCompute访问身份不可修改为个人账号，否则生产环境的数据安全将无法得到保障。

- 工作空间列表页面显示的是生产环境工作空间的名称（默认任务运行在开发环境工作空间）。
- 工作空间成员被授予的系统角色，仅对开发环境工作空间生效。子账号被授予的系统角色，在生产环境工作空间不生效，您需要在数据地图进行申请，详情请参见[#unique_28](#)。
- 子账号默认无权限对生产环境工作空间进行操作。如果您使用子账号访问生产环境工作空间下的其他产品，例如DataHub投递数据至MaxCompute表等，需要重新给予账号授权。您可以通过MaxCompute客户端进行授权，也可以在数据地图页面申请权限，详情请参见[#unique_29](#)和[#unique_28](#)。

项目关系

通过MaxCompute或DataWorks官网产品页进入[管理控制台](#)创建的项目时，有两种选择：

- 简单模式的项目实际上是创建了关联绑定好的一个MaxCompute Project和一个DataWorks项目空间（工作空间）。同时在MaxCompute的Project里创建对应的几个Role，具体Role权限请参见成员角色权限关系。
- 标准模式的项目实际上是创建了关联绑定好的一个开发（dev）MaxCompute Project、一个生产（prod）MaxCompute Project来同时对应一个DataWorks项目空间（工作空间）。同时在MaxCompute的Project里创建对应的几个Role，具体Role权限请参见成员角色权限关系。

用户和权限关系图

一个DataWorks项目空间绑定一个MaxCompute project，此时您可以根据DataWorks项目管理 > 工作空间配置 > 访问身份这个属性设置决定DataWorks其他项目成员是否拥有MaxCompute project的权限。

访问身份分为个人账号和计算引擎指定账号（系统账号），具体用户和权限对应关系如下图所示。

对于标准模式，一个DataWorks项目空间绑定两个MaxCompute project：一个开发项目和一个生产项目。DataWorks其他项目成员根据成员角色拥有MaxCompute开发project对应的role权限，但没有MaxCompute生产project的权限。MaxCompute任务需要通过发布流程发布到生产project后，以owner账号提交到MaxCompute执行。

6 版本历史

本文将为您及时同步DataWorks的版本更新。

DataWorks V2.0发布

发布版本：DataWorks V2.0

- 发布时间：2018年7月25日
- 发布范围：华东2区域
- 发布内容：DataWorks V2.0在DataWorks V1.0的基础上，新增业务流程、组件的概念，并完善数据开发体系。DataWorks V2.0支持双工作空间开发，将开发和生产隔离，保证数据研发规范，减少错误代码的出现。

已升级DataWorks V2.0区域

目前所有区域均升级至DataWorks V2.0版本。

DataWorks V2.0版本操作差异

- DataWorks升级至V2.0版本后：此前版本中已存在的补数据实例将不再显示。如果您需要检查在升级前夕，生成的补数据实例状态，可以进入数据管理 > 数据表管理 > 分区信息页面，查看是否有新的时间分区产生。如果在预期时间内，仍未显示补数据实例的业务日期，建议您重新创建补数据实例。

如果您通过上述方法，无法判断补数据实例是否成功创建，请您提交工单进行咨询。

- 您可以参见DataWorks V2.0系列详解视频，对DataWorks V2.0各功能模块的特性进行深入学习：
 - [DataWorks V2.0版本概述与最佳实践](#)
 - [DataWorks V2.0前生后世](#)
 - [DataWorks V2.0常见问题与难点分析](#)
 - [DataWorks V2.0数据开发功能与用法解析](#)
 - [DataWorks V2.0数据集成简介与最佳实践](#)
 - [DataWorks V2.0智能监控简介与最佳实践](#)
 - [DataWorks V2.0数据服务功能及用法解析](#)
 - [DataWorks V2.0数据质量简介及最佳实践](#)
 - [DataWorks V2.0数据安全简介与最佳实践](#)
 - [Function Studio简介与使用指导](#)

DataWorks V2.0版本更新一览表

DataWorks V2.0提升了整体视觉交互和数据开发模块的使用体验。新增智能监控、数据保护伞、数据质量和数据服务四大模块。

如果您想快速了解新老版本的差异，请参见下表的版本更新说明和讲解视频[DataWorks V2.0版本概述与最佳实践](#)。

模块名称	子模块	对比项	DataWorksV1.0	DataWorksV2.0	改进效果
MaxCompute项目	项目管理模式	管理方式	一个 DataWorks 项目对应一个 MaxCompute 项目。	引入标准模式的概念。一个 DataWorks 工作空间对应两个 MaxCompute 项目，分别是开发环境和生产环境，详情请参见 简单模式和标准模式的区别 。	隔离风险，进一步保障生产环境代码的稳定性。

模块名称	子模块	对比项	DataWorksV1.0	DataWorksV2.0	改进效果
数据开发	任务开发	整体功能	进行单任务、工作流的代码编写、周期调度配置，完成后可提交运维中心自动调度。	<ul style="list-style-type: none"> · 更名：名称更改为数据研发。 · 新增：解决方案、业务流程的概念。 · 删减：删除工作流的概念。 · 优化：SQL编辑器更加智能，任务周期配置、依赖配置更加开放。 	<ul style="list-style-type: none"> · SQL编辑器：提供更加人性化、沉浸式的SQL开发体验。 · 任务管理：业务流程、解决方案让复杂的开发任务管理更加简单。
		SQL研发	以单个任务或工作流的形式，在页面上编写SQL代码并进行试运行。	<p>提供更加智能的SQL编辑器以满足您的需求，可以实现：代码高亮、格式化、智能补齐、错误提示和表结构显示等人性化功能。</p> <p>同时，您可以在编辑器右侧结构标签内，以图形化的形式直观地对SQL内部结构进行梳理。</p>	<ul style="list-style-type: none"> · 任务调度：更开放的调度系统，轻松应对更加复杂的业务场景。 · 其他功能：新增功能可以优化长久以来的用户痛点，从细节上照顾用户的体验。

模块名称	子模块	对比项	DataWorksV1.0	DataWorksV2.0	改进效果
		节点配置	通过单个节点、工作流模式组合业务代码。	引入业务流程的概念替代工作流。您可以在业务流程中自由组合任务，并根据需求，将不同的资源归入业务流程中进行管理（所有任务、表、资源、函数都必须从属于一个业务流程）。您可以通过解决方案，对业务流程进行整合，统一管理业务关联性较强的流程。	
		周期配置	工作流整体周期配置会影响单个任务的周期配置。	无法对组合后的节点进行整体周期配置，但可以对所有节点单独进行配置，且调度周期类型不受上下游节点影响。	
		依赖属性	工作流之间的依赖局限于整体。	不同业务流程中的任务节点可以相互依赖，不需要依赖业务流程整体。	
	脚本开发	整体功能	作为周期任务的补充，通常用于非周期的临时数据处理，如临时表的增删改等，因此不包含周期属性和依赖关系。	功能相同，更名为临时查询。	
	手动任务	整体功能	创建的所有节点都需要手动触发，无法通过调度执行。	功能相同，更名为手动业务流程。	

模块名称	子模块	对比项	DataWorksV1.0	DataWorksV2.0	改进效果
	资源管理	整体功能	作为一个单独标签管理该 MaxCompute 项目中的所有资源，包括 jar/file/archive。	作为业务流程中的一个子标签，您可以按需加入该业务流程中涉及到的资源，同时可以建立多层文件夹进行管理。	
	函数管理	整体功能	作为一个单独标签，管理该 MaxCompute SQL 编辑时需要的系统函数和自定义函数。	既可以作为一个单独标签存在并管理所有函数，也可以作为业务流程中的一个子标签，仅管理该业务流程中所需使用的函数。	
	表查询	整体功能	展示该 MaxCompute 项目下的所有表，同时能够预览内容、引用表名。	相同。	
	表管理（新增）	整体功能	无	供开发者管理自己的表。可以进行生命周期设置、表管理（包括修改表的类目、描述、字段、分区等）、表隐藏/取消隐藏和表删除等操作。	
	临时查询	整体功能	无	用于在本地测试代码的实际情况与期望值是否相符，无提交、发布、设置调度参数功能。	
	组件管理（新增）	整体功能	无	将大量相似并可复用的 SQL 代码抽象为 SQL 代码块或节点任务，可供您自由地配置输入输出参数，并将其运用到各个不同的实际业务中。	

模块名称	子模块	对比项	DataWorksV1.0	DataWorksV2.0	改进效果
	运行历史 (新增)	整体功能	无	展示最近三天内, 在本地运行过的所有任务记录。同时可查看任务运行结果, 并提供简单的过滤功能。	
	结果筛选 (新增)	整体功能	无	页面SQL运行的结果集成了Excel组件, 让您在页面输出结果后, 通过简单的过滤、筛选、排序得到所需结果。	
	回收站 (新增)	整体功能	无	用于防止您误删任务所带来的业务损失。您可以在回收站中查看当前工作空间下所有删除的节点, 并提供恢复功能。	
	代码全局搜索	整体功能	无	您可以输入不完整字符串来查找包含该字符串的MaxCompute SQL、Shell、数据同步任务, 快速定位自己所需查看或操作的任务。	
	发布功能	整体功能	DataWorksV1.0 标准模式项目下保留发布功能。	更名为项目克隆。仅简单模式的工作空间具备主动将任务克隆至其他工作空间的功能。	

模块名称	子模块	对比项	DataWorksV1.0	DataWorksV2.0	改进效果
运维中心	任务列表	功能	根据节点类型、名称和责任人搜索任务。	增加通过业务流程、解决方案、基线名称来搜索任务的功能。	从业务视角来对任务进行运维，匹配了任务开发界面的新功能。
	任务运维	功能	根据节点类型、名称、责任人、业务日期和运行日期搜索任务。	增加通过业务流程、解决方案、基线名称搜索任务的功能。	
	报警	功能	通过出错、完成和未完成等事件，作为监控报警的依据。	集成基线监控、事件报警、自定义报警三大功能，构建更加智能、完备的报警体系。	
智能监控（新增）	<p>智能监控（Alarm）是DataWorks任务运行的监控及分析系统。根据监控规则和任务运行情况，智能监控决策是否报警、何时报警、如何报警以及给谁报警。智能监控会自动选择最合理的报警时间、报警方式和报警对象。</p>				赋予用户在云上一站式完成数据开发、数据（安全）治理、数据分享的闭环体验。
数据质量DQC（新增）	<p>数据质量，是支持多种异构数据源的质量校验、通知、管理服务的一站式平台。</p> <p>数据质量以数据集（DataSet）为监控对象，目前支持MaxCompute数据表和DataHub实时数据流的监控。当离线MaxCompute数据发生变化时，数据质量会对数据进行校验并阻塞生产链路，避免问题数据污染扩散。同时，数据质量提供了历史校验结果的管理，以便您对数据质量分析和定级。</p>				
数据服务（新增）	<p>数据服务提供了快速将数据表生成数据API的能力，支持您将现有的API快速注册到数据服务平台以统一管理和发布。同时，数据服务已与API网关（API Gateway）打通，支持将API服务一键发布至API网关。数据服务与API网关一道为您提供安全稳定、低成本、易上手的数据开放共享服务。</p>				
数据保护伞（新增）	<p>数据保护伞为MaxCompute提供数据资产识别、敏感数据发现、数据分级分类、脱敏、访问监控、风险发现预警与审计的能力。</p>				

7 文档最新动态

本文为您介绍DataWorks 2.0版本文档更新的最新动态，基于此您可以及时了解DataWorks 2.0版本新增特性及功能变更。

时间	特性	类别	描述	适用客户	产品文档
2019.08.23	新增插件	新增功能	为您介绍AnalyticDB for MySQL 3.0插件支持的数据类型、字段映射和数据源等参数及配置示例。	所有	#unique_38 和 #unique_39
2019.8.22	新增SLA	新增	为您介绍DataWorks服务等级协议（SLA）。	所有	#unique_40
2019.08.21	新增查看账单详情	体验优化	为您介绍如何查看DataWorks按量计费商品中，各计费项的计费明细账单。	所有	#unique_41
2019.08.14	新增使用教程	体验优化	窃电用户自动识别将为您介绍如何通过DataWorks配合机器学习的方式，实现窃电用户的自动识别，保障用户的安全用电。	所有	窃电用户自动识别教程
2019.07.08	新增使用教程	体验优化	数据质量是数据分析结论有效性和准确性的基础。本教程为您介绍如何通过数据质量风险监控，保证数据的完整性、准确性。	所有	数据质量保障教程
2019.07.04	新增节点	新功能	您可以在Dataworks中新建AnalyticDB for PostgreSQL节点，构建在线ETL数据处理流程。	所有	#unique_44
2019.07.04	新增节点	新功能	您可以在Dataworks中新建AnalyticDB for MySQL节点，构建在线ETL数据处理流程。	所有	#unique_45
2019.07.04	新增节点	新功能	您可以在Dataworks中新建Data Lake Analytics节点，构建在线ETL数据处理流程。	所有	#unique_46

时间	特性	类别	描述	适用客户	产品文档
2019.06.21	新增使用教程	体验优化	本教程助您了解并实操阿里云的大数据产品，完成在线运营分析平台的搭建。	所有	搭建互联网在线运营分析平台教程
2019.05.21	新增数据地图	功能升级	介绍数据管理升级到数据地图的进展和计划。	所有	数据地图
2019.05.20	新增产品定价	商业化	DataWorks将于6月份开始正式实施商业化，当前计费价格仅供参考，具体日期将另行通知。	所有	产品定价
2019.05.08	新增研发规范	体验优化	介绍数据仓库研发规范的阶段规划、角色职责和整体流程。	数据研发者和管理者	研发规范
2019.04.16	新增节点组	新功能	您可在数据开发页面新增和引用节点组。	所有	#unique_51
2019.04.15	新增遍历节点	新功能	通过for-each节点实现循环2次，每次循环中把当前的循环次数打印出来的需求。	所有	#unique_52
2019.04.10	新增ODPS Spark节点	新功能	DataWorks提供ODPS Spark任务类型，您可进行新建和配置。	所有	#unique_53
2019.04.08	新增最佳实践	体验优化	通过Shell节点结合自定义资源组的方法来实现邮件外发的需求。	所有	#unique_54
2019.03.29	新增最佳实践	体验优化	每5分钟抽取一次数据，待每天00:00的同步任务抽取完成后，对当天共288次同步任务抽取的所有数据进行计算。	所有	#unique_55
2019.03.20	新增循环节点	新功能	您可在do-while节点中定义相互依赖的任务，任务中包含一个名为end的循环判断节点。Dataworks会不断重复执行这一批任务，直到循环判断节点end把判断结果置为false，Dataworks才会退出整个循环。	所有	#unique_56

时间	特性	类别	描述	适用客户	产品文档
2019.02.27	新增 Stream Studio	新功能	Stream Studio基于阿里云实时计算引擎构建，集DAG和SQL两种开发模式为一体，是DataWorks旗下的一站式实时计算开发平台。	所有	#unique_57
2019.02.20	新增自定义节点	新功能	DataStudio不仅支持原生的ODPS SQL、Shell等系统节点，也支持自定义节点来满足需求。	项目owner和项目经理	自定义节点
2019.02.20	新增安全中心	新功能	您可在安全中心模块进行组织内全局数据表权限的查看、表权限管理、数据表权限申请/审批等操作。	所有	安全中心
2019.01.28	新增最佳实践	体验优化	使用数据服务开发数据API，快速在DataV中调用API并展现数据分析结果。	所有	#unique_60
2019.01.15	新增最佳实践	体验优化	新增分支、赋值、归并节点最佳实践。	所有	#unique_61
2019.01.04	新增节点类型	新功能	目前DataWorks已支持赋值节点。	所有	#unique_62
2019.01.04	新增节点类型	新功能	目前DataWorks已支持分支节点。	所有	#unique_63
2019.01.04	新增节点类型	新功能	目前DataWorks已支持归并节点。	所有	#unique_64
2019.01.02	RDS白名单更新	特性变更	各Region新增IP，美东地区新增安全组。	所有	#unique_65 , #unique_66
2018.12.13	RDS白名单更新	特性变更	雅加达region新增IP: 11.194.49.0/24,11.200.93.0/24,11.200.95.0/24,11.200.97.0/24,100.64.0.0/8,149.129.228.0/24,10.143.32.0/24	所有	#unique_65

时间	特性	类别	描述	适用客户	产品文档
2018.12.03	RDS白名单更新	特性变更	悉尼region新增IP: 11.192.100.0/24,11.192.134.0/24, 11.192.135.0/24,11.192.184.0/24,11.192.99.0/24,100.64.0.0/8,47.91.49.0/24,47.91.50.0/24	所有	#unique_65
2018.11.28	新增最佳实践	体验优化	目前DataWorks已支持Python UDF。	所有	使用MaxCompute分析IP来源最佳实践
2018.11.28	RDS白名单更新	特性变更	杭州region新增IP: 114.55.197.0/24,11.197.246.0/24, 11.197.247.0/24 日本region新增IP: 11.199.250.0/24,47.91.27.0/24	所有	#unique_65
2018.11.27	新增Function Studio功能	新功能	Function Studio是一个面向函数开发场景的WEB项目代码编辑开发工具。	所有	#unique_67 、 #unique_68 、 Function Studio快速开始
2018.11.16	支持新类型数据源及读写插件	新功能	支持POLARDB数据源、POLARDB Writer、POLARDB Reader、HybridDB for MySQL数据源、HybridDB for MySQL Writer、HybridDB for MySQL Reader、AnalyticDB for PostgreSQL数据源、AnalyticDB for PostgreSQL Writer、AnalyticDB for PostgreSQL Reader。	所有	#unique_70 、 #unique_71 、 #un
2018.11.12	支持实时转实例功能	新功能	使用实时转实例功能，您可以在代码编辑完成发布后即时生成实例。	所有	#unique_79
2018.10.25	新增批量上云功能	新功能	为您介绍如何批量添加数据源，并将其中的数据收集到云端。	所有	#unique_80