



数据湖工厂

常见问题

文档版本 02

发布日期 2019-05-24

华为技术有限公司



版权所有 © 华为技术有限公司 2020。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默示的声明或保证。

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。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编： 518129

网址： <https://www.huawei.com>

客户服务邮箱： support@huawei.com

客户服务电话： 4008302118

目录

1 DLF 是什么?	1
2 DLF 可以做什么?	2
3 什么是作业?	3
4 DLF 可以创建多少个作业?	4
5 DLF 的作业配额是否可以变更?	5
6 DLF 是否提供作业模板?	6
7 什么是节点?	7
8 DLF 作业中的节点数是否有限制?	8
9 作业关联的 CDM 集群删除后, 如何快速修复?	9
10 作业的计划时间和开始时间相差大, 是什么原因?	10
11 如何修改消息的语言?	11
12 作业失败怎么处理?	12

1 DLF 是什么?

数据湖工厂（Data Lake Factory，简称数据开发模块）是华为云的大数据平台产品，提供一站式的大数据协同开发环境、全托管的大数据调度能力，同时支持批处理、实时Pipeline编排、监控、运维，图形化编程（图形化编程可以减少80% SQL），帮助用户快速构建大数据处理中心。

DLF的版本管理功能还在规划中，后续会开放此功能，敬请期待。

2 DLF 可以做什么？

使用数据开发模块，您可以实现以下功能：

- 数据开发模块支持多种类型的数据连接和数据表，可以根据需求设计和构建数据模型。
- 数据开发模块提供20多种异构数据源之间的数据迁移，实现多数据源集成到数据连接。
- 数据开发模块支持在线脚本编辑调试、拖拽式作业开发等，实现数据开发工作。
- 数据开发模块提供强大的作业调度与灵活的监控告警，轻松管理数据作业运维。

3 什么是作业?

作业由一个或多个节点组成，共同执行以完成对数据的一系列操作。

4 DLF 可以创建多少个作业?

目前默认每个用户最多可以创建1000个作业。

5 DLF 的作业配额是否可以变更?

可以，系统支持用户根据实际需求调整最大配额。

如有需求，请[提交工单](#)进行申请。

6 DLF 是否提供作业模板?

提供，您可以通过数据开发模块提供的基础模板进行服务体验。

7 什么是节点?

节点定义对数据执行的操作。

例如，使用“数据迁移”节点将数据从OBS导入到MRS Hive。

8 DLF 作业中的节点数是否有限制?

有限制，每个作业最多只能包含100个节点。

9 作业关联的 CDM 集群删除后，如何快速修复？

问题

如果DLF作业关联的CDM集群被删除，那么我们如何快速修复，尽可能减小对DLF作业的影响。

回答

CDM集群被删除后，DLF作业中的关联信息会保留原配置。用户只需在CDM中新建同名集群和作业，DLF作业将使用新的同名CDM集群和作业，同时提示用户原CDM集群和作业将被替代。

限制条件：

该功能于1.7.3版本（上线时间：2018-10-24）实现，此前已创建的DLF作业如需使用该功能，请重新保存作业。

10 作业的计划时间和开始时间相差大，是什么原因？

问题

在作业监控页面查看作业运行记录时，发现作业的计划时间和开始时间相差较大，导致这个问题的原因是什么？

图 10-1 问题示例图



The screenshot shows a table with the following data:

状态	调度方式	计划时间	开始时间	结束时间	运行时间 (min)	操作
运行成功	正常调度	2018/10/09 16:50:00 GMT +08...	2018/10/09 17:50:08 GMT +08...	2018/10/09 17:59:28 GMT +08...	9.3	停止 重试 继续执行 强制成功

回答

计划时间是作业预期开始执行的时间，即用户为作业配置的调度计划。开始时间是作业实际开始执行的时间。

在DLF中，单个作业最多允许5个实例并行执行，如果作业实际执行时间大于作业配置的调度周期，会导致后面批次的作业实例堆积，从而出现上述问题。

出现上述问题时，请检查作业配置的调度周期是否小于作业实际执行所需要的时间，根据实际情况调整作业的调度计划。

11 如何修改消息的语言？

问题

用户在DLF的通知管理中配置了邮件/短信通知，当作业运行异常时，向指定人员发送消息。目前，消息的内容是中文的，如何修改为英文的？

回答

消息语言的修改方法如图11-1所示：

图 11-1 修改通知消息语言



1. 在华为云控制台右上方，单击用户名，进入账号中心页面。
2. 单击“首选项设置”。
3. 选择“消息接收设置”页签，配置消息语言为“English”，同时还可以修改消息的时区。

12 作业失败怎么处理？

问题

当相互依赖的几个作业，在作业调度过程中某个作业执行失败，是否会影响后续作业？这时该如何处理？

回答

会影响后续作业，后续作业可能会挂起，继续执行或终止执行。

图 12-1 作业依赖关系

* 依赖的作业失败后，当前作业处理策略

挂起 继续执行 终止执行

这时请勿停止作业，您可以将失败的作业实例进行重跑，或者将异常的实例停止再重跑。失败实例成功后，后续作业会继续正常运行。如果不通过DLF，手动将作业实例中的业务场景处理后，可以强制成功作业实例，后续作业也会继续正常运行。