



数据接入服务

常见问题

文档版本 01

发布日期 2019-06-21

华为技术有限公司



版权所有 © 华为技术有限公司 2019。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。

目录

1 什么是 DIS?	4
2 什么是分区?	5
3 DIS 主要应用于哪些场景?	6
4 DIS 有哪些特点和优势?	7
5 DIS 如何发送和接收数据?	8
6 如何开通 DIS 通道?	9
7 数据存储于 DIS 和转储其他资源有什么区别?	10
8 DIS 有哪些模块及各模块功能?	12
9 如何校验软件包完整性?	13

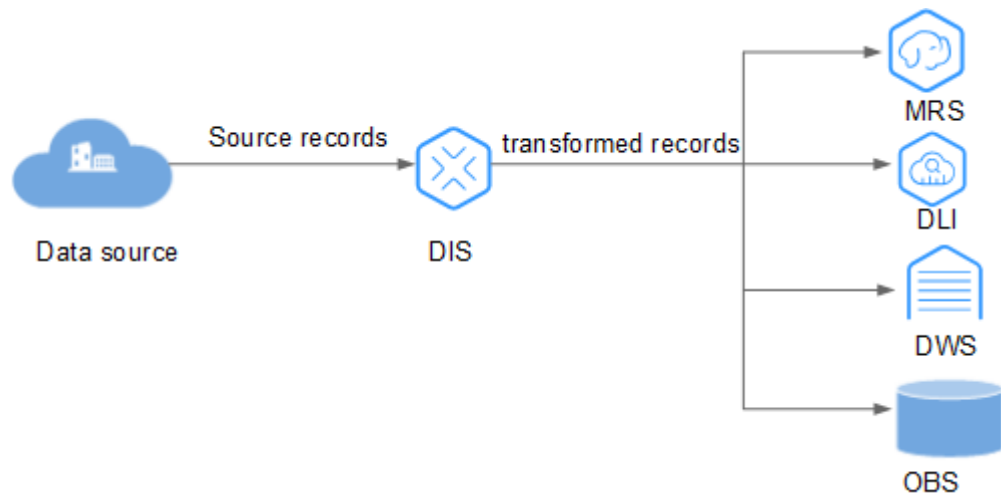
1 什么是 DIS?

数据接入服务（Data Ingestion Service）为处理或分析流数据的自定义应用程序构建数据流管道，主要解决云服务外的数据实时传输到云服务内的问题。数据接入服务每小时可从数十万种数据源（如IoT数据采集、日志和定位追踪事件、网站点击流、社交媒体源等）中连续捕获、传送和存储数TB数据。

数据流向

- DIS实时从多种数据源采集数据。
- DIS连续传输数据，自动将数据传输至MRS，DLI，DWS和OBS等服务做计算，分析和存储。

图 1-1 数据流向



2 什么是分区?

分区 (Partition) 是DIS数据通道的基本吞吐量单位。创建通道时，将指定所需的分区数量。

目前每个租户默认Partition配额范围为1~50个，租户可以根据需要配置Partition个数。

3 DIS 主要应用于哪些场景？

DIS对于从数据生产者快速移出数据，然后进行持续处理非常有用。以下是使用DIS的典型场景：

- **加速日志和数据传送获取：**您无需等待批量处理数据，而是让数据生产者在生成数据后立即输入DIS数据通道，防止因数据生产者出现故障导致的数据损失。例如，系统和应用程序日志可以持续添加到数据通道并可在数秒内进行处理。
- **实时指标和报告：**实时从DIS数据通道数据提取指标并生成报告。例如，数据接入服务应用程序可以处理系统和应用程序日志的指标和报告，因为数据被流入而不是等待收到批量数据。
- **实时数据分析：**通过数据接入服务，可以运行实时通道数据分析。例如，可以通过API把数据实时添加到DIS数据通道中，并让您的DIS应用程序实时运行分析，从而在数分钟内从数据中获得重要见解，而无需数小时或数天时间。
- **复杂的数据通道处理：**您可以创建DIS应用程序和数据通道的Directed Acyclic Graphs (DAG)。在这一情景中，一个或多个DIS应用程序可将数据添加到一个DIS数据通道进行进一步处理，以便于进行通道处理器的后续阶段。

4 DIS 有哪些特点和优势？

- 无限扩展：DIS数据通道的吞吐量每小时可从数MB扩展到数TB，PUT记录每秒钟可从数千次扩展到数百万。
- 易于使用：您可以在几秒钟内创建DIS数据通道，轻松的将数据放入通道中，并构建用于数据处理的应用程序。
- 成本低廉：DIS没有前期成本，您只需要为实际使用的资源付费即可。
- 并行处理：DIS可让您用多个应用程序同时处理同一个数据通道。例如，您可以让一个应用程序运行实时分析，让其他应用程序从同一个DIS数据通道中将数据发送至对象存储服务（Object Storage Service，简称OBS）。
- 安全可靠：DIS可将数据保留 $N*24$ 小时， N 的取值为1~7的整数，以防数据在应用程序故障、个别机器故障或设施故障时丢失。

5 DIS 如何发送和接收数据？

- 步骤1** 开通DIS通道，在IAM（用户认证中心）中获取账号的AK/SK。
 - 步骤2** 在<https://dis-publish.obs-website.cn-north-1.myhwclouds.com/>中下载“dis-sdk-X.X.X.zip”压缩包并解压缩。
 - 步骤3** 建立工程，配置用户AK/SK、endpoint、projectId、region、通道名称、分区数量等。
 - 步骤4** 配置完成后运行程序即可发送数据。
 - 步骤5** 建立工程，配置用户AK/SK、endpoint、project、region、通道名称、partitionId和startingSequenceNumber。
 - 步骤6** 配置完成后运行程序即可接收数据。
- 结束

6 如何开通 DIS 通道？

用户开通DIS通道的方法请参见《数据接入服务 用户指南》。

7 数据存储 在 DIS 和转储其他资源有什么区别？

开通DIS通道时需要选择“转储服务类型”。具体区别如表7-1所示。

- 选择“OBS”表示存储在DIS中，并周期性导入对象存储服务（Object Storage Service，简称OBS）。
- 选择“MRS”表示存储在DIS中，并周期性导入MapReduce服务（MRS）集群的HDFS中。
- 选择“DLI”表示存储在DIS中，并周期性导入DLI。
- 选择“DWS”表示存储在DIS中，并周期性导入数据仓库服务（DWS）中。
- 选择“CloudTable”表示存储在DIS中，并实时导入CloudTable集群的HBase表或OpenTSDB表中。

表 7-1 DIS 和转储其他资源区别

DIS存储	OBS存储	MRS存储	DLI存储	DWS存储	CloudTable存储
DIS服务自带。	需要另外申请。	需要另外申请。	需要另外申请。	需要另外申请。	需要另外申请。
无需另外付费。	需要根据OBS收费标准另外付费。	需要根据MRS和OBS收费标准另外付费。	需要根据DLI和OBS收费标准另外付费。	需要根据DWS和OBS收费标准另外付费。	需要根据CloudTable收费标准另外付费。
临时存储（最长保留168小时）。	数据可长期存储在OBS中，具体保存时长根据用户购买的OBS服务时长决定。	数据可长期存储在MRS中，具体保存时长根据用户购买的MRS服务时长决定。	数据可长期存储在DLI中，具体保存时长根据用户购买的DLI服务时长决定。	数据可长期存储在DWS中，具体保存时长根据用户购买的DWS服务时长决定。	数据可长期存储在CloudTable中，具体保存时长根据用户购买的CloudTable服务时长决定。

DIS存储	OBS存储	MRS存储	DLI存储	DWS存储	CloudTable存储
只存储在DIS中。	存储在DIS中，并周期性导入OBS。	存储在DIS中，并周期性导入MRS集群的HDFS中。 说明 导入MRS集群前临时存储在OBS，待转储MRS完成后删除OBS上的临时存储文件。	存储在DIS中，并周期性导入DLI。 说明 导入DLI前临时存储在OBS，待转储DLI完成后删除OBS上的临时存储文件。	存储在DIS中，并周期性导入DWS。 说明 导入DWS前临时存储在OBS，待转储DWS完成后删除OBS上的临时存储文件。	存储在DIS中，实时导入CloudTable集群的HBase表或OpenTSDB表中。

8 DIS 有哪些模块及各模块功能？

- 服务控制面
 - 完成服务的开通、删除、配置操作，并将用户信息同步到数据面。
 - 完成数据面资源的申请与自动部署。
- 服务数据面
 - 接收用户发送数据的请求，对已鉴权的数据接收并存储。
 - 接收用户获取数据的请求，在鉴权后输出对应的用户数据。
 - 按时老化存储在系统中的用户数据。
 - 根据用户配置，将用户数据存储到对象存储服务（Object Storage Service，简称OBS）。
- 服务维护
 - 负责服务的安装、升级。
 - 负责服务的配置、巡检、日志收集与分析、运行监控。
 - 负责服务工单处理。
- 用户SDK
 - 提供Java接口，供用户上传与下载数据。
 - 提供数据加密功能。

9 如何校验软件包完整性?

获取DIS SDK软件包及校验文件后，可以在Linux系统上按如下步骤对软件包的完整性进行校验。

前提条件

- 已获取“PuTTY”工具。
- 已获取“WinSCP”工具。

操作步骤

步骤1 使用“WinSCP”工具将“huaweicloud-sdk-dis-x.x.x.zip”上传至Linux系统任一目录。



x.x.x表示DIS SDK包的版本号。

步骤2 使用“PuTTY”工具登录Linux系统，进入到“huaweicloud-sdk-dis-x.x.x.zip”所在目录，执行如下命令，获取DIS SDK压缩包的校验码。

```
sha256sum huaweicloud-sdk-dis-x.x.x.zip
```

显示类似如下校验码：

```
# sha256sum dis-sdk-x.x.x.zip  
8be2c937e8d78b1a9b99777cee4e7131f8bf231de3f839cf214e7c5b5ba3c088 huaweicloud-sdk-dis-x.x.x.zip
```

步骤3 打开DIS SDK的校验文件“huaweicloud-sdk-dis-x.x.x.zip.sha256sum”与上一步骤中获取的校验码进行对比。

- 一致，说明从获取的DIS SDK压缩包没被篡改。
- 不一致，说明DIS SDK压缩包被篡改，需要重新获取。

----**结束**