

数据集成 常见问题



腾讯云

【 版权声明 】

©2013–2025 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【 商标声明 】



及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【 服务声明 】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【 联系我们 】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或 95716。

文档目录

常见问题

实时同步常见问题

离线同步常见问题

其他常见问题

常见问题

实时同步常见问题

最近更新时间：2024-06-18 14:47:41

MySQL 类

1. MySql serverid 冲突

错误信息：

```
com.github.shyiko.mysql.binlog.network.ServerException: A slave with the
same server_uuid/server_id as this slave has connected to the master。
```

解决办法：目前已经优化增加随机生成 serverid，之前的任务中如果在 mysql 高级参数中显示指定了 server-id 建议删除，因为可能多个任务使用了相同的数据源，并且 server-id 设置的相同导致冲突。

2. binlog 文件找不到：

错误信息：

```
Caused by: org.apache.kafka.connect.errors.ConnectException: The
connector is trying to read binlog starting at GTIDs xxx and binlog file
'binlog.xxx', pos=xxx, skipping 4 events plus 1 rows, but this is no
longer available on the server. Reconfigure the connector to use a
snapshot when needed。
```

错误原因：

作业正在读取的 binlog 文件在 MySQL 服务器已经被清理时，会产生报错。导致 Binlog 清理的原因较多，可能是 Binlog 保留时间设置的过短；或者作业处理的速度追不上 Binlog 产生的速度，超过了 MySQL Binlog 文件的最大保留时间，MySQL 服务器上的 Binlog 文件被清理，导致正在读的 Binlog 位点变得无效。

解决办法：如果作业处理速度无法追上 Binlog 产生速度，可以考虑增加 Binlog 的保留时间也可以优化作业减轻反压来加速 source 消费。如果作业状态没有异常，可能是数据库发生了其他操作导致 Binlog 被清理，从而无法访问，需要结合 MySQL 数据库侧的信息来确定 Binlog 被清理的原因。

3. mysql 连接被重置

错误信息：

```
EventDataDeserializationException: Failed to deserialize data of EventHeaderV4 .... Caused by: java.net.SocketException: Connection reset。
```

错误原因:

1. 网络问题。
2. 作业存在反压, 导致 source 无法读取数据, binlog client 空闲, 如果 binlog 连接在超时后仍然空闲 mysql 服务器会断开空闲的连接。

解决方法:

1. 如果是网络问题, 可以调大 mysql 网络参数 `set global slave_net_timeout = 120;` (默认30s) `set global thread_pool_idle_timeout = 120。`
2. 如果是作业反压导致, 可以通过调节作业减轻反压, 比如增加并行度, 提升写入速度, 提升 taskmanager 内存减少 gc。

4. Mysql2dlc 任务 JobManager Oom

错误信息:

```
op_type, meta.database_name, meta.op_ts]) -> Calc(select=[CAST(meta.database_name) AS database_wedata_di, CAST(meta.op_type) AS type_wedata_di, CAST(meta.op_ts) AS ts_wedata_di, CAST(PROCTIME_MATERIALIZE(PROCTIME())) AS processing_time_wedata_di, CAST(id) AS id, maoyanid AS maoyanid, sourcecode, other_id, comment_id, username, avatar_url, comment_time, content, score, digg_count, CAST(createtime) AS createtime, CAST(modifytime) AS modifytime, entity_type]) -> NotNullEnforcer(fields=[id])] ERROR org.apache.flink.runtime.util.FatalExceptionHandler - FATAL: Thread 'SourceCoordinator-Source: TableSourceScan(table=[[default_catalog, default_database, table_9806_146847]], fields=[id, maoyanid, sourcecode, other_id, comment_id, username, avatar_url, comment_time, content, score, digg_count, createtime, modifytime, entity_type, meta.table_name, meta.op_type, meta.database_name, meta.op_ts]) -> Calc(select=[CAST(meta.database_name) AS database_wedata_di, CAST(meta.table_name) AS table_wedata_di, CAST(meta.op_type) AS type_wedata_di, CAST(meta.op_ts) AS ts_wedata_di, CAST(PROCTIME_MATERIALIZE(PROCTIME())) AS processing_time_wedata_di, CAST(id) AS id, maoyanid AS maoyanid, sourcecode, other_id, comment_id, username, avatar_url, comment_time, content, score, digg_count, CAST(createtime) AS createtime, CAST(modifytime) AS modifytime, entity_type]) -> NotNullEnforcer(fields=[id])' produced an uncaught exception. Stopping the process...
java.lang.OutOfMemoryError: Java heap space
    at java.util.Arrays.copyOf(Arrays.java:3236) ~[?:1.8.0_332]
    at java.io.ByteArrayOutputStream.toByteArray(ByteArrayOutputStream.java:191) ~[?:1.8.0_332]
    at org.apache.flink.runtime.source.coordinator.SourceCoordinator.writeCheckpointBytes(SourceCoordinator.java:404) ~[flink-dist_2.11-1.13.6.jar:1.13.6]
    at org.apache.flink.runtime.source.coordinator.SourceCoordinator.toBytes(SourceCoordinator.java:385) ~[flink-dist_2.11-1.13.6.jar:1.13.6]
    at org.apache.flink.runtime.source.coordinator.SourceCoordinator.lambda$checkpointCoordinator$5(SourceCoordinator.java:259) ~[flink-dist_2.11-1.13.6.jar:1.13.6]
6)
    at org.apache.flink.runtime.source.coordinator.SourceCoordinator$$Lambda$1715/581528218.run(Unknown Source) ~[?:?]
    at org.apache.flink.runtime.source.coordinator.SourceCoordinator.lambda$runInEventLoop$8(SourceCoordinator.java:344) ~[flink-dist_2.11-1.13.6.jar:1.13.6]
    at org.apache.flink.runtime.source.coordinator.SourceCoordinator$$Lambda$1519/398851494.run(Unknown Source) ~[?:?]
    at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149) ~[?:1.8.0_332]
    at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624) ~[?:1.8.0_332]
```

错误原因和解决方法:

1. 用户数据量比较大, 可以调大 jobmanager CU 数, 使用 mysql 高级参数 `scan.incremental.snapshot.chunk.size` 调大 chunk size 大小, 默认是8096。
2. 用户数据量不大, 但是主键最大值-最小值的差值却很大, 导致使用均分 chunk 的策略时划分很多 chunk, 修改分布因子, 让用户数据走非均匀的数据切分逻辑, `split-key.even-distribution.factor.upper-`

bound=5.0d，默认值分布因子已经修改为10.0d。

5. 用户的 binlog 数据格式不对，导致 debezium 解析异常

错误信息：

```
ERROR io.debezium.connector.mysql.MySqlStreamingChangeEventSource [] -
Error during binlog processing. Last offset stored = null, binlog reader
near position = mysql-bin.000044/211839464.
2023-02-20 21:37:28.480 [blc-172.17.48.3:3306] ERROR
io.debezium.pipeline.ErrorHandler [] - Producer failure
io.debezium.DebeziumException: Error processing binlog event.
```

解决方法：

修改 binlog_row_image=full 后需要重启数据库。

6. 是否支持 gh-ost?

支持，不会迁移 Online DDL 变更产生的临时表数据，只迁移源库使用 gh-ost 执行的原始 DDL 数据，同时您可以使用默认的或者自行配置 gh-ost 影子表和无用表的正则表达式。

DLC 类

1. 同步增量数据（changlog 数据）到 dlc 时报错

错误信息：

```
2022-07-12 21:03:57.161 [IcebergFilesCommitter -> Sink: IcebergSink HYBRIS.xny_sl03_pro.subs_crm_clue (1/1)#34] WARN org.apache.flink.runtime.taskmanager.Task [] -
IcebergFilesCommitter -> Sink: IcebergSink HYBRIS.xny_sl03_pro.subs_crm_clue (1/1)#34 (ced203d857ca39a5a1eb207be88110b5) switched from INITIALIZING to FAILED with failure cause:
java.lang.IllegalArgumentException: Cannot write delete files in a v1 table
    at org.apache.iceberg.ManifestFiles.writeDeleteManifest(ManifestFiles.java:154)
    at org.apache.iceberg.SnapshotProducer.newDeleteManifestWriter(SnapshotProducer.java:374)
    at org.apache.iceberg.MergingSnapshotProducer.lambda$newDeleteFilesAsManifests$10(MergingSnapshotProducer.java:681)
    at java.util.HashMap.forEach(HashMap.java:1289)
    at org.apache.iceberg.MergingSnapshotProducer.newDeleteFilesAsManifests(MergingSnapshotProducer.java:678)
    at org.apache.iceberg.MergingSnapshotProducer.prepareDeleteManifests(MergingSnapshotProducer.java:664)
    at org.apache.iceberg.MergingSnapshotProducer.apply(MergingSnapshotProducer.java:533)
    at org.apache.iceberg.SnapshotProducer.apply(SnapshotProducer.java:164)
    at org.apache.iceberg.SnapshotProducer.lambda$commit$2(SnapshotProducer.java:283)
    at org.apache.iceberg.util.Tasks$Builder.runTaskWithRetry(Tasks.java:404)
    at org.apache.iceberg.util.Tasks$Builder.runSingleThreaded(Tasks.java:214)
    at org.apache.iceberg.util.Tasks$Builder.run(Tasks.java:198)
    at org.apache.iceberg.util.Tasks$Builder.run(Tasks.java:190)
    at org.apache.iceberg.SnapshotProducer.commit(SnapshotProducer.java:282)
    at org.apache.iceberg.flink.sink.IcebergFilesCommitter.commitOperation(IcebergFilesCommitter.java:312)
    at org.apache.iceberg.flink.sink.IcebergFilesCommitter.commitDeltaTxn(IcebergFilesCommitter.java:299)
    at org.apache.iceberg.flink.sink.IcebergFilesCommitter.commitUpToCheckpoint(IcebergFilesCommitter.java:218)
    at org.apache.iceberg.flink.sink.IcebergFilesCommitter.initializeState(IcebergFilesCommitter.java:153)
    at org.apache.flink.streaming.api.operators.StreamOperatorStateHandler.initializeOperatorState(StreamOperatorStateHandler.java:118)
    at org.apache.flink.streaming.api.operators.AbstractStreamOperator.initializeState(AbstractStreamOperator.java:290)
    at org.apache.flink.streaming.runtime.tasks.OperatorChain.initializeStateAndOpenOperators(OperatorChain.java:441)
    at org.apache.flink.streaming.runtime.tasks.StreamTask.restoreGates(StreamTask.java:585)
    at org.apache.flink.streaming.runtime.tasks.StreamTaskActionExecutor$1.call(StreamTaskActionExecutor.java:55)
```

问题原因：

dlc 的表是 v1表不支持 changlog 数据。

解决方法：

1. 修改 dlc 表为 v2表和开启 upsert 支持。
 2. 修改表支持 upsert: ALTER TABLE tblname SET TBLPROPERTIES ('write.upsert.enabled'='true')。
 3. 修改表为 v2表: ALTER TABLE tblname SET TBLPROPERTIES ('format-version'='2')。
- 显示表设置的属性是否成功 show tblproperties tblname。

2. Cannot write incompatible dataset to table with schema

错误详情：

```
Caused by: java.lang.IllegalArgumentException: Cannot write incompatible
dataset to table with schema:
* mobile should be required, but is optional
    at
org.apache.iceberg.types.TypeUtil.checkSchemaCompatibility(TypeUtil.java
:364)
    at
org.apache.iceberg.types.TypeUtil.validateWriteSchema(TypeUtil.java:323)
```

错误原因：

用户建 dlc 表时，字段设置了 NOT NULL 约束。

解决办法：

建表时不要设置 not null 约束。

3. 同步 mysql2dlc，报数组越界

问题详情：

```
java.lang.ArrayIndexOutOfBoundsException: 1
    at
org.apache.flink.table.data.binary.BinarySegmentUtils.getLongMultiSegmen
ts(BinarySegmentUtils.java:736) ~[flink-table-blink_2.11-
1.13.6.jar:1.13.6]
    at
org.apache.flink.table.data.binary.BinarySegmentUtils.getLong(BinarySegm
```

```
entUtils.java:726) ~[flink-table-blink_2.11-1.13.6.jar:1.13.6]
    at
    org.apache.flink.table.data.binary.BinarySegmentUtils.readTimestampData(
    BinarySegmentUtils.java:1022) ~[flink-table-blink_2.11-
    1.13.6.jar:1.13.6]
    at
    org.apache.flink.table.data.binary.BinaryRowData.getTimestamp(BinaryRowD
    ata.java:356) ~[flink-table-blink_2.11-1.13.6.jar:1.13.6]
    at
    org.apache.flink.table.data.RowData.lambda$createFieldGetter$39385f9c$1(
    RowData.java:260) ~[flink-table-blink_2.11-1.13.6.jar:1.13.6]
```

错误原因：

时间字段做主键导致的问题。

解决方法：

1. 不用时间字段做主键。
2. 用户还是需要时间字段保证唯一性，建议用户在 dlc 新增一个冗余字符串字段，将上游的时间字段使用函数转换后映射到冗余字段。

4. DLC 任务报不是一个 iceberg 表

错误信息：

```
[flink-dist_2.11-1.13.6.jar:1.13.6]
... 10 more
Caused by: org.apache.iceberg.exceptions.NoSuchIcebergTableException: Not an iceberg table: HYBRIS.origindb.
movieunifiedorder01_unified_order_sharding_maoyan_order_v2 (type=null)
    at org.apache.iceberg.exceptions.NoSuchIcebergTableException.check(NoSuchIcebergTableException.java:33) ~
[pool-1-thread-1-1678868648050-prod-sort-connector-iceberg-dlc-2.6.29-1-1.4.0.jar:1.4.0]
```

错误原因：

1. 可以让用户在 dlc 执行语句查看表到底是什么类型。
2. 语句：desc formatted 表名。
3. 可以看表类型。

解决办法：

选择正确的引擎建 iceberg 类型的表。

5. flink sql 字段顺序跟 dlc 目标表字段不一致导致报错

解决方法：

不修改任务表字段顺序。

Doris 类

1. Doris 规格如何选型及调优？

请参考 [Doris 资源规格选型及调优建议](#)。

2. 导入任务过多，新导入任务提交报错 “current running txns on db xxx is xx, larger than limit xx” ？

调整 fe 参数：max_running_txn_num_per_db，默认100，可适当调大，建议控制在500以内。

3. 导入频率太快出现 err=[E-235] 错误？

- 参数调优建议：可通过适当调大 max_tablet_version_num 参数暂时解决，此参数默认200，建议控制在2000以内。
- 业务调优建议：降低导入频率才能根本解决这个问题。

4. 导入文件过大，被参数限制。报错 “The size of this batch exceed the max size” ？

调整 be 参数：streaming_load_max_mb，建议超过需要导入的文件大小。

5. 导入数据报错：“[-238]” ？

- 原因：-238 错误通常出现在同一批导入数据量过大的情况，从而导致某一个 tablet 的 Segment 文件过多（由）。
- 参数调优建议：可适当调大BE参数max_segment_num_per_rowset，此参数默认值200，可按倍数调大（如400、800），建议控制在2000以内；
- 业务调优建议：建议减少一批次导入的数据量。

6. 导入失败，报错：“too many filtered rows xxx, "ErrorURL": "或 Insert has filtered data in strict mode, tracking url=xxxx.” ？

原因：表的 schema、分区等与导入的数据不匹配。可在 CDW Studio 或客户端执行 doris 命令查看具体原因：show load warnings on `<tracking url>`，<tracking url> 即为报错信息中返回的 error url。

离线同步常见问题

最近更新时间：2024-03-28 17:06:01

网络类

确认 ccn 产品跨区域网络打通需要公网带宽

数据集成资源组需要和客户的数据库在同一个 VPC 网络下，通过内网访问。如果集成资源组和数据库在不同区域的 VPC 下，需要通过云联网 CCN 打通网络。

云联网 CCN 为了方便测试，默认放通了10Kb的流量，只适合小数据量的同步和网络测试。如果用于性能测试和生产任务，请购买带宽，避免数据同步受到网络带宽限制。

←

ccn-xxxxxT8i3 详情

关联实例

监控

带宽管理

路由表

i

为了便于测试连通性，

所有地域间10Kbps以下带宽免费

购买带宽

带宽ID	地域A ▾	地域B ▾	供应
------	-------	-------	----

Hive 类

Hive On CHDFS 表写入报错

错误信息：
Permission denied: No access rules matched

```
at com.tencent.teg.scheduler.runner.runners.WeDataDiRunner.execute(WeDataDiRunner.java:101)
at com.tencent.cloud.wedata.runner.common.api.AbstractTaskRunner.executeWithCommonLog(AbstractTaskRunner.java:508)
at com.tencent.cloud.wedata.runner.common.api.AbstractTaskRunner.startWork(AbstractTaskRunner.java:546)
at com.tencent.teg.scheduler.runner.runners.WeDataDiRunner.main(WeDataDiRunner.java:75)
[2022-12-29 01:00:52]-[ERROR] 获取FileSystem时发生网络IO异常, 请检查您的网络是否正常!HDFS地址: [message:defaultFS=ofs://[redacted],myqcloud.com]
[2022-12-29 01:00:52]-[ERROR] Exception when job run: com.tencent.wedata.common.exception.wedataException: Code:[HiveWriter-06], Description:[与HDFS建立连接时出现IO异常.] - IOException: Permission denied: No access rules matched
at com.tencent.wedata.common.exception.wedataException.aswedataException(wedataException.java:26)
at com.tencent.wedata.plugin.writer.hivewriter.HdfsHelper.getFileSystem(HdfsHelper.java:368)
at com.tencent.wedata.plugin.writer.hivewriter.HiveWriter$Job.init(HiveWriter.java:194)
at com.tencent.wedata.core.job.JobContainer.initJobWriter(JobContainer.java:717)
at com.tencent.wedata.core.job.JobContainer.init(JobContainer.java:307)
at com.tencent.wedata.core.job.JobContainer.start(JobContainer.java:114)
at com.tencent.wedata.core.Engine.start(Engine.java:232)
at com.tencent.wedata.core.Engine.entry(Engine.java:122)
at com.tencent.teg.scheduler.runner.runners.wedata.wedataHelper.startwedata(wedataHelper.java:126)
at com.tencent.teg.scheduler.runner.runners.WeDataDiRunner.execute(WeDataDiRunner.java:161)
at com.tencent.cloud.wedata.runner.common.api.AbstractTaskRunner.executeWithCommonLog(AbstractTaskRunner.java:508)
at com.tencent.cloud.wedata.runner.common.api.AbstractTaskRunner.startWork(AbstractTaskRunner.java:546)
at com.tencent.teg.scheduler.runner.runners.WeDataDiRunner.main(WeDataDiRunner.java:75)

[2022-12-29 01:00:52]-[INFO] Total 0 records, 0 bytes | Speed 0B/s, 0 records/s | Error 0 records, 0 bytes | All Task WaitWriterTime 0.000s | All Task WaitReaderTime 0.000s | Percenta
[2022-12-29 01:00:52]-[INFO]

该任务最可能的错误原因是:
wedatException: Code:[HiveWriter-06], Description:[与HDFS建立连接时出现IO异常.] - IOException: Permission denied: No access rules matched
[2022-12-29 01:00:52]-[ERROR] Task execution failed
[2022-12-29 01:00:52]-[INFO] commitTask success.
[2022-12-29 01:00:52]-[INFO] taskId: <[redacted]>
[2022-12-29 01:00:52]-[INFO] taskState: <FAILED>
```

错误原因:

数据集成资源组是全托管资源组，资源组的出口网段非客户的VPC内网网段，需要对数据集成资源组放开安全组。

解决方法:

1. 请在 CHDFS 挂载点，对集成资源组的 CHDFS 权限 id 进行授权。



云 HDFS

基础配置 高级配置 安全配置 挂载点

添加挂载点

输入挂载点名称搜索

名称	挂载地址	VPC 权限组	状态	操作
wedata	f84r...dfs.ap-beijing.myqcloud.com	VpcId: vpc-... 权限组Id: ag-... VpcId: 无权限 权限组Id: ag-...	可用	编辑 删除 绑定权限组 解绑权限组

数据集成资源组的chdfs权限id

2. 打开页面：API 3.0 Explorer 填入参数，其中 MountPointId 为客户 CHDFS 挂载点，AccessGroupIds.N 为 数据集成资源组的 CHDFS 权限 id（注意，不同区域的权限id不同）。

云 API

信息中心

API 库

API Explorer

API Doctor(使用诊断)

错误码中心

API Explorer

云 HDFS (CHDFS)

搜索接口, 支持中英文搜索

其他接口

挂载点相关接口

修改挂载点属性

解绑权限组列表

查看挂载点列表

查看挂载点详细信息

删除挂载点

创建挂载点

绑定权限组列表

生命周期规则相关接口

文件系统相关接口

权限规则相关接口

权限组相关接口

回热任务相关接口

AssociateAccessGroups

chdfs 2020-11-12 查看API文档

点赞 吐槽

代码生成 在线调用 签名串生成 参数说明 问题反馈 查看文档 数据模拟

在线调用

tccli chdfs AssociateAccessGroups --cli-unfold-argument --region ap-beijing --MountPoint

Golang Python Java C++ Node.js PHP .Net

调试SDK示例代码

package main

import (

"fmt"

"github.com/tencentcloud/tencentcloud-sdk-go/tencentcloud/common"

"github.com/tencentcloud/tencentcloud-sdk-go/tencentcloud/common/errors"

"github.com/tencentcloud/tencentcloud-sdk-go/tencentcloud/common/profile"

chdfs "github.com/tencentcloud/tencentcloud-sdk-go/tencentcloud/chdfs/v20201112"

)

func main() {

// 实例化一个认证对象, 入参需要传入腾讯云账户 SecretId 和 SecretKey, 此处还需注意密钥对的保

// 代码泄露可能会导致 SecretId 和 SecretKey 泄露, 并威胁账号下所有资源的安全性。以下代码示

// 例仅供参考, 请谨慎使用。详细请参见 https://cloud.tencent.com/document/product/1278/85305

// 密钥可前往官网控制台 https://console.cloud.tencent.com/cam/capi 进行获取

credential := common.NewCredential(

"SecretId",

"SecretKey",

)

// 实例化一个client选项, 可选的, 没有特殊需求可以跳过

cpf := profile.NewClientProfile()

cpf.HttpProfile.Endpoint = "chdfs.tencentcloudapi.com"

// 实例化要请求产品的client对象, clientProfile是可选的

client, _ := chdfs.NewClient(credential, "ap-beijing", cpf)

// 实例化一个请求对象, 每个接口都会对应一个request对象

request := chdfs.NewAssociateAccessGroupsRequest()

}

更多选项

输入参数

Region

华北地区(北京) ap-beijing 集成资源组的区域

参数输入方式

表单 JSON 参数推荐

MountPointId

18411uennm33le5 CHDFS挂载点

AccessGroupIds.N

1 ydc-g3lnoft6 集成资源组的CHDFS权限id

添加

说明：

数据集成资源组 CHDFS 权限 id 和区域对照表：

区域	数据集成资源组 CHDFS 权限 id
北京	ag-wgbku4no
广州	ag-x1bhppnr
上海	ag-xnjfr9d3
新加坡	ag-phxtv0ah
美国硅谷	ag-tgwl8bca
弗吉尼亚	ag-esxpwxjn

其他常见问题

最近更新时间：2024-12-10 11:41:52

1. 数据集成支持哪些数据源？

数据集成（DataInLong）支持多种云上及自建数据源，例如：

- 关系型数据库：Mysql、PostgreSQL、Oracle、SQL Server、IBM DB2、达梦 DM、SAP HANA 等。
- 大数据存储：Hive、HDFS、HBase、Kudu、Clickhouse、DLC、Impala、Gbase、Tbase 等。
- 半结构化：FTP、SFTP、COS。
- NoSQL：Redis、Elasticsearch。
- 消息队列：Kafka 等。

2. 数据集成支持哪些同步场景？

数据集成（DataInLong）核心支持离线和实时数据同步场景。

- 离线场景支持周期性数据采集；
- 实时场景基于数据库日志/消息队列监控数据变更，实时刷新最新数据。

3. 支持哪些数据同步方向？

数据集成采用星型结构数据模型，针对支持数据源类型可灵活任意搭配来源和去向。

4. 支持何种数据对象同步方式？

数据集成支持单库单表、分库分表、批量多库多表同步方式。