

分布式关系型数据库数据 迁移接口技术要求

Technical requirements for data migration interfaces of distributed
relational databases

征求意见稿

XXXX - XX - XX 发布

XXXX - XX - XX 实施

目 次

前 言 II

引 言 1

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 1

5 接口功能 1

 5.1 概述 1

 5.2 接口访问控制 2

 5.3 存量数据迁移 2

 5.4 增量数据迁移 6

 5.1 错误码规则 9

 5.2 JDBC 类型映射说明 9

6 增量数据 JSON 模型 9

 6.1 DML 输出 JSON 模型实例 9

参 考 文 献 12

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由北京信息化协会提出并归口。

本文件起草单位：。

本文件主要起草人：。

引 言

近年来，分布式关系型数据库在金融、通信、能源等多个行业得到了大量应用。制定分布式关系型数据库数据迁移接口标准对于降低用户数据被特定的分布式关系型数据库产品绑定的风险、推动分布式关系型数据库技术的发展、促进产业生态建设等方面都具有重要的意义。通过规范源端数据库数据迁移接口的功能、协议与数据模型等内容，便于目标端数据库或第三方迁移工具调用，将数据安全、可靠、高效地迁移并加载到目标端数据库。

规范和统一接口：通过制定统一的分布式关系型数据库数据迁移接口标准，可以确保不同数据库产品之间的可迁移性。

分布式关系型数据库数据迁移接口技术要求

1 范围

本文件规定了分布式关系型数据库在数据迁移接口的功能、协议、数据模型等内容。

本文件适用于分布式关系型数据库在数据迁移过程中的数据的迁移（不涉及表结构、触发器、存储过程、函数等对象），接口由被迁移端数据库实现。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 32633—2016 分布式关系数据库服务接口规范。

3 术语和定义

GB/T 37740—2019界定的以及下列术语和定义适用于本文件。

3.1

数据迁移 DATA MIGRATION

利用数据库系统或第三方提供的各类服务，实现应用数据从源数据库端向目标数据库端转移的过程。

3.2

返回值 RETURN VALUE

说明接口函数执行后返回结果的可能值。

3.3

功能 FUNCTIONALITY

说明接口函数的用途。

4 缩略语

JDBC Java 数据库连接 (Java Database Connectivity)

JSON JavaScript 对象表示法 (JavaScript Object Notation)

HTTPS 安全超文本传输协议 (HyperText Transfer Protocol Secure)

5 接口功能

5.1 概述

分布式关系型数据库数据迁移接口主要包括接口访问控制、存量数据迁移、增量数据迁移等功能。

5.2 接口访问控制功能

5.2.1 接口功能描述

接口访问控制功能用于对访问接口的用户的身份鉴别。

接口采用? ? ? ? 实现，输入参数包括用户名和密码；返回参数包括.....。

5.2.2 输入参数

参数	类型	是否必填	描述
USERNAME	STRING	是	鉴权登录用户名
PASSWORD	STRING	是	登录用户密码，BASE64 加密

5.2.3 返回参数

参数名	类型	描述
STATUS	INT	任务状态：1-成功，0-失败
ERROR_CODE	STRING	错误码
ERROR_MESSAGE	STRING	错误描述

5.3 存量数据迁移

存量任务迁移包括如下两个步骤：

- a) 数据迁移组件向数据库发起任务创建请求。
- b) 数据迁移组件查询任务状态。

5.3.1 存量任务创建功能

5.3.1.1 接口功能描述

存量任务创建接口用于数据迁移组件向数据库发起任务创建请求。

接口采用? ? ? 实现，输入待迁移数据范围、文件导出目录路径、压缩算法、数据块大小等参数，并返回对应参数。

5.3.1.2 输入参数

参数名	类型	必填	描述
DB_TABLES	ARRAY	是	待迁移库表清单，支持通配符（如 "DB.*"），格式： ["DB1.SCHEMA1.TABLE1", "DB2.*"]
STORAGE_PATH	STRING	是	输出目录路径，支持 NFS/S3 等协议（示例：S3://BUCKET/PATH/）
COMPRESSION	STRING	否	压缩算法（默认不压缩，可选：ZSTD、SNAPPY、GZIP、LZ4）
CHUNK_SIZE_MB	INT64	否	数据分块大小（默认 128MB，范围 16-1024）

5.3.1.3 返回参数

5.3.1.4 处理要求

参数名	类型	归属上级	描述
STATUS	INT	/	任务状态：1-成功，0-失败
TASK_ID	STRING	/	任务 ID：FULL_TASK_{DATE}_{ID}
ERROR_CODE	STRING	/	错误码
ERROR_MESSAGE	STRING	/	错误描述

TASK_ID：任务ID，FULL_TASK_{DATE}_{ID}，确保任务编码的唯一性。

5.3.2 存量任务状态查询接口

5.3.2.1 接口描述

存量任务状态查询接口用于数据迁移组件查询任务状态,以确定当前存量数据迁移进行状态与进度,通过输入任务ID，返回存量数据迁移任务归属的数据库、表、已导出表元数据信息文件名等信息。

5.3.2.2 输入参数

参数名	类型	必填	描述
TASK_ID	STRING	是	任务 ID：FULL_TASK_{DATE}_{ID}

5.3.2.3 返回参数

参数名	类型	归属上级	描述
TASK_ID	STRING	/	任务 ID：FULL_TASK_{DATE}_{ID}
TASK_STATUS	STRING	/	RUNNING/FINISHED
TOTAL_TIME_SEC	INT64	/	任务运行总耗时（秒），运行中的任务返回已运行时长
TABLE_FILES	LIST	/	已完成导出的表文件信息
DB_NAME	STRING	TABLE_FILES	该表归属的 DB
TABLE_NAME	STRING	TABLE_FILES	表名
TABLE_RECORDS	DOUBLE	TABLE_FILES	已导出的表的记录数
TABLE_META_FILE	STRING	TABLE_FILES	已导出表的元数据信息的文件名称
FILE_NUMBER	INT64	TABLE_FILES	表导出的文件数量
STORAGE_PATH	STRING	TABLE_FILES	基础存储路径
FILE_NAMES	LIST	TABLE_FILES	导出的文件名列表
FILE_NAME	STRING	FILE_NAMES	导出的文件名

ERROR_CODE	STRING	/	错误码 (如 ERR_EXPORT_TIMEOUT)
ERROR_MESSAGE	STRING	/	错误描述

5.3.2.4 处理要求

规范:

- 1. 文件命名规则: {TASK_ID}/{DB}_{TBL}_{CHUNKID:08D}_{CRC32:08X}.{EXT}
TASK_ID: 任务ID, FULL_TASK_{DATE}_{ID}
DB: 数据库名
TBL: 表名
CHUNKID:08D: 文件大小超过128MB后, 进行分块, 对分块编排序号(8位零填充)
CRC32:08X: 文件内容的CRC32校验 (8位16进制)
EXT: 扩展名: DAT/ZST/SNAPPY/GZ

- 2. 数据文件格式规则:
文件是标准json, 结构如下:

```
{
  {
    "column_name1" : " data1" ,
    "column_name2" : " data2" ,
    "column_name3" : " data3"

  },
  {
    "column_name1" : " data3" ,
    "column_name2" : " data4" ,
    "column_name3" : " data5"
  }
}
```

- 3. 元数据文件格式规则:
文件是标准json, 结构如下:

```
{
  "schema": "schema1",           // 模式名
  "table_name": "table1",        // 表名
  "total_columns": 5,            // 总列数
  "primary_key": ["id"],         // 主键列名列表
  "columns": [                   // 列定义数组
    {
```

```

    "column_name": "id",          // 列名
    "jdbc_type": "BIGINT",       // JDBC 标准类型
    "precision": 20,             // 精度 (数字类型)
    "scale": 0,                  // 小数位数 (数字类型)
    "is_nullable": false,        // 是否允许 NULL
  },
  {
    "column_name": "name",
    "jdbc_type": "VARCHAR",
    "precision": 255,            // 字符长度
    "is_nullable": true
  },
  {
    "column_name": "salary",
    "jdbc_type": "DECIMAL",
    "precision": 10,             // 总位数
    "scale": 2,                  // 小数位
  },
  {
    "column_name": "created_at",
    "jdbc_type": "TIMESTAMP",
    "precision": 6               // 时间精度 (微秒)
  },
  {
    "column_name": "metadata",
    "jdbc_type": "CLOB",         // 大文本类型
    "precision": null            // 不适用
  }
],
"table_comment": "用户信息表", // 表注释 (可选)
"version": "1.0"                // 元数据版本
}

```

5.3.3 停止存量迁移接口

5.3.3.1 接口描述

停止存量迁移接口用于数据迁移组件发起停止存量数据迁移任务请求，通过输入任务ID，获取任务状态等信息。

5.3.3.2 输入参数

参数名	类型	必填	描述
TASK_ID	STRING	是	任务 ID: FULL_TASK_{DATE}_{ID}

5.3.3.3 返回参数

参数名	类型	归属上级	描述
STATUS	INT	/	任务状态: 1-成功, 0-失败
TASK_ID	STRING	/	任务 ID: FULL_TASK_{DATE}_{ID}
END_POSITION	STRING	/	截止时间戳 (ISO8601 格式, STOP_POSITION_TYPE=TIMESTAMP 时必填)
ERROR_CODE	STRING	/	错误码
ERROR_MESSAGE	STRING	/	错误描述

5.4 增量数据迁移

增量任务迁移包括三个步骤:

1. 数据迁移组件向数据库发起任务创建请求。
2. 数据迁移组件查询任务状态。
3. 数据迁移组件发起停止任务请求。

5.4.1 增量任务创建接口

5.4.1.1 接口描述

增量任务创建接口，用于数据迁移组件向数据库发起任务创建请求，输入待迁移库表清单、文件导出路径、数据块大小等参数，返回对应信息。

5.4.1.2 输入参数

参数名	类型	必填	描述
DB_TABLES	ARRAY	是	待迁移库表清单，支持通配符（如 "DB.*"），格式: ["DB1.SCHEMA1.TABLE1", "DB2.*"]
STORAGE_PATH	STRING	是	输出目录路径，支持 NFS/S3 等协议（示例: S3://BUCKET/PATH/）
CHUNK_SIZE_MB	INT64	否	数据分块大小（默认 128MB，范围 16-1024）
START_POSITION_TYPE	STRING	是	起始位置类型: LSN/TIMESTAMP

START_POSITION	STRING	是	物理日志序列号（十六进制字符串， START_POSITION_TYPE=LSN 时必填） 起始时间戳（ISO8601 格式， START_POSITION_TYPE=TIMESTAMP 时必填）
----------------	--------	---	---

5.4.1.3 返回参数

参数名	类型	归属上级	描述
STATUS	INT	/	任务状态：1-成功，0-失败
TASK_ID	STRING	/	任务 ID：INC_TASK_{DATE}_{ID}
ERROR_CODE	STRING	/	错误码（如 ERR_EXPORT_TIMEOUT）
ERROR_MESSAGE	STRING	/	错误描述

5.4.1.4 处理要求

- 1. 当出现增量数据变更时，输出增量JSON到指定目录。
- 2. 增量数据通过APPEND方式写入目标文件块，增量数据采用JSON格式，具体的格式参考附录 “6 增量数据JSON模型” 。
- 3. 文件命名规则： {TASK_ID}/INC_{DB}_{CHUNKID:08D}.JSON
TASK_ID：任务ID， INC_TASK_{DATE}_{ID}
DB：数据库名
CHUNKID:08D：分块序号(8位零填充)

5.4.2 增量任务状态查询接口

5.4.2.1 接口描述

增量任务状态查询接口用于数据迁移组件查询任务状态，通过输入任务ID，获取任务归属数据库名、表名、已导出表记录数、表导出文件数量等信息。

5.4.2.2 输入参数

参数名	类型	必填	描述
TASK_ID	STRING	是	任务 ID：INC_TASK_{DATE}_{ID}

5.4.2.3 返回参数

参数名	类型	归属上级	描述
TASK_ID	STRING	/	任务 ID：INC_TASK_{DATE}_{ID}
TASK_STATUS	STRING	/	RUNNING/STOPED
TOTAL_TIME_SEC	INT64	/	任务运行总耗时（秒），运行中的任务返回已运行时长
TABLE_FILES	LIST	/	已完成导出的表文件信息

DB_NAME	STRING	TABLE_FILES	该表归属的 DB
TABLE_NAME	STRING	TABLE_FILES	表名
TABLE_RECORDS	DOUBLE	TABLE_FILES	已导出的表的记录数
TABLE_META_FILE	STRING	TABLE_FILES	已导出表的元数据信息的文件名称
FILE_NUMBER	INT64	TABLE_FILES	表导出的文件数量
STORAGE_PATH	STRING	TABLE_FILES	基础存储路径
FILE_NAMES	LIST	TABLE_FILES	导出的文件名列表
FILE_NAME	STRING	FILE_NAMES	导出的文件名
ERROR_CODE	STRING	/	错误码（如 ERR_EXPORT_TIMEOUT）
ERROR_MESSAGE	STRING	/	错误描述

5.4.2.4 处理要求

数据规范：

1. 文件命名规则：{TASK_ID}/{DB}_{TBL}_{CHUNKID:08D}_{CRC32:08X}.{EXT}
TASK_ID：任务ID，INC_TASK_{DATE}_{ID}
DB：数据库名
TBL：表名
CHUNKID:08D：分块序号(8位零填充)
CRC32:08X：文件内容的CRC32校验（8位16进制）
EXT：扩展名: DAT/ZST/SNAPPY/GZ
2. 数据文件格式规则：见增量数据JSON模型。

5.4.3 停止增量迁移接口

5.4.3.1 接口描述

停止增量迁移接口用于数据迁移组件发起停止任务请求，通过输入任务ID，获取任务状态，物理日志序列号等信息。

5.4.3.2 输入参数

参数名	类型	必填	描述
TASK_ID	STRING	是	任务 ID: INC_TASK_{DATE}_{ID}

5.4.3.3 返回参数

参数名	类型	归属上级	描述
STATUS	INT	/	任务状态：1-成功，0-失败
TASK_ID	STRING	/	任务 ID: INC_TASK_{DATE}_{ID}

END_POSITION	STRING	/	物理日志序列号（十六进制字符串， START_POSITION_TYPE=LSN 时必填） 截止时间戳（ISO8601 格式， STOP_POSITION_TYPE=TIMESTAMP 时必填）
ERROR_CODE	STRING	/	错误码
ERROR_MESSAGE	STRING	/	错误描述

5.4.3.4 处理要求

增量迁移是持续进行的动作，需要主动停止增量迁移动作。

5.5 错误码规则

错误码前缀	分类	示例	描述示例
ERR_	致命错误	ERR_EXPORT_TIMEOUT	导出操作超时
WARN_	警告	WARN_TYPE_CONVERSION	类型不一致，隐式转换

5.6 数据类型映射说明

数据库类型	数据类型	精度要求	示例值
INT	INTEGER	显示精度 (e.g., 11)	"precision":11
BIGINT	BIGINT	显示精度 (e.g., 20)	"precision":20
VARCHAR	VARCHAR	字符长度	"precision":255
DECIMAL	DECIMAL	总位数和小数位	"precision":10, "scale":2
DATETIME	TIMESTAMP	时间精度 (0-6)	"precision":0 (秒) "precision":3 (毫秒) "precision":6 (微秒)
BLOB	BLOB	字节长度 (可选)	"precision":4096
BOOLEAN	BOOLEAN	不需要	省略 precision

6 增量数据 JSON 模型

6.1 DML 输出 JSON 模型实例

6.1.1 INSERT

```
{  
  "TYPE": "DML",
```

```
"OPERATION": "INSERT",
"SCHEMA": "SCHEMA1",
"TABLE": "TABLE1",
"TIMESTAMP": "2025-07-01T10:30:45.123Z",
"AFTER": {
  "ID": 1001,
  "NAME": "NEW USER",
  "EMAIL": "NEW@EXAMPLE.COM"
}
```

6.1.2 UPDATE

```
{
  "TYPE": "DML",
  "OPERATION": "UPDATE",
  "SCHEMA": "SCHEMA1",
  "TABLE": "TABLE1",
  "TIMESTAMP": "2025-07-01T10:31:22.456Z",
  "BEFORE": {
    "ID": 1001,
    "NAME": "NEW USER",
    "EMAIL": "NEW@EXAMPLE.COM"
  },
  "AFTER": {
    "ID": 1001,
    "NAME": "UPDATED USER",
    "EMAIL": "UPDATED@EXAMPLE.COM"
  }
}
```

6.1.3 DELETE

```
{
  "TYPE": "DML",
  "OPERATION": "DELETE",
  "SCHEMA": "SCHEMA2",
```

```
"TABLE": "TABLE2",  
"TIMESTAMP": "2025-07-01T10:32:15.789Z",  
"BEFORE": {  
  "ID": 2001,  
  "PRODUCT": "LAPTOP",  
  "PRICE": 1299.99  
}  
}
```

参 考 文 献

- [1] GB/T 37740—2019 信息技术 云计算 云平台间应用和数据迁移指南
- [2] GB/T 34079.2-2021 基于云计算的电子政务公共平台服务规范 第2部分：应用部署和数据迁移