

TiDB数据迁移

1.Dumpling

2.lightning

- 1.介绍
 - 全量数据高速导入到tidb集群中
- 2.原理过程
 - 导入模式
 - 对功能有影响
 - 建立schema和表
 - 分割表
 - 读取sql dump
 - 写入本地临时存储文件
 - 导入数据到tikv集群
 - 校验和分析
 - 普通模式
- 3.适用场景
 - 大量新的数据需要快速导入到tidb数据库中
 - 400gb/h - 500gb/h数据导入
- 4.支持的数据格式
 - dumpling
 - csv
 - 亚马逊s3云盘或者本地磁盘
- 5.后端
 - local-backend
 - 4.0以后
 - 速度快 500gb/h
 - importer-backend
 - 过一手
 - tidb-backend
 - 导入线上生产环境，且表中有数据
- 6.部署需求
 - 32+cpu
 - 20gb+内存
 - ssd硬盘
 - 万兆网卡
 - 单独部署
- 7.安装
 - tiup install lightning
 - tidb-toolkit包
- 8.配置
 - 配置文件
- 9.启动
 - nohup ./tidb-lightning --config tidb-lightning.toml & nohup.out &
 - 成功返回 exit
 - 断点续传
 - checkpoint
 - 启动断点续传功能
 - driver
 - 断点续传的存储
 - 控制
 - checkpoint-error-destroy
 - 失败的表从头开始
 - checkpoint-error-ignore
 - 忽略之前的错误进行继续导入
 - checkpoint-remove
 - 无论是否有出错，把表的断点清楚
- 10.数据过滤
 - f 参数
 - 正则匹配
 - toml配置文件
 - mydumper项加入filter
- 11.web界面
 - 启动
 - 命令行
 - ./tidb-lightning --server-mode --status-addr :8289
 - 配置文件
 - lightning模块
 - 提交任务
 - 查看任务进度
 - 管理任务

- 1.介绍
 - 把mysql/tidb的数据导出为csv或者sql格式
 - 逻辑上的全量备份或者导出
- 2.架构与特点
 - 支持多种数据形式
 - tidb server上
 - 逻辑导出
 - 全新的表过滤和数据过滤
 - 支持导出到亚马逊s3云盘
 - 保持一致性
 - 针对tidb做出优化
- 3.使用场景
 - 小于50GB
 - 数据量较小
 - 可以进行异构
 - 对于导出效率要求不高
- 4.不适用
 - 导出sst文件
 - 目前只能全量
 - 增量备份
 - 数据量较大场景
- 5.部署安装
 - tiup install dumpling
 - tidb-toolkit
- 6.所需权限
 - select
 - reload
 - lock tables
 - replication client
- 7.使用方式
 - 1.导出sql文件

```
dumpling \
-u root\
-P 4000\
-h 127.0.0.1\
--filetype sql \
--threads 32\
-o /tmp/test\
-r 200000\
-F 256MB
```
 - 2.导出csv
 - filetype改成csv
 - 3.数据筛选
 - 过滤
 - where "id<100"
 - 只能导出CSV
 - filter 正则匹配表
 - 和where区别是一个是部分数据 一个是整个的
 - B 数据库
 - T table
 - 4.导出数据的格式
 - metadata
 - 包含文件导出的起始时间，
 - master binary log的位置
 - 创建schema的sql语句
 - schema-cerate.sql
 - 创建table的sql文件
 - schema.table-schema.sql
 - 创建表数据的文件
 - schema.table.001.csv|sql
 - 5.导出数据的一致性
 - 参数--consistency <consistency level>
 - flush
 - 可读不可写
 - FTWRL
 - 开始执行flush table with read lock
 - snapshot
 - 默认
 - ./dumpling --snapshot 'date|timestamp'
 - lock
 - 只锁需要备份的表
 - none
 - 不保持一致性
 - auto
 - 会自动去选择一致性
 - tidb 默认snapshpt
 - 6.Dumpling性能优化
 - 1. -t 指定导出的线程数
 - 2. -r 导出文件记录的最大行数