

## 组网及说明

DE E0107环境, Hbase作为主题库, MPP1 (高可用) 作为应用库, 业务要求从主题库Hbase将数据根据业务规则抽取并加载到MPP1中。如直接配置表抽取-表加载任务, 由于该方式加载使用insert动作, 当MPP1目的表数据量大了后效率极低。而goload加载数据的效率很高, 所以需要组合gpload工具和达梦的etl, 达梦会根据前台页面的相关治理按钮生成etl任务, 做到自动完成goload的动作, 同时也支持增量加载。

## 配置步骤

1. 配置达梦etl服务到MPP1 Master的无密码登录信任关系
2. 表抽取。达梦etl工具根据前台调用将数据源表生成表抽取任务, 将抽取数据以csv格式存放在指定的路径下
3. 文件同步。达梦etl工具配置文件同步任务, 将抽取出来的csv文件同步到MPP1 Master服务器, 并删除etl服务器本地的文件
4. Master节点配置goload的yaml配置文件  
如源目表字段数目和顺序一致, 则可省略columns部分定义, 否则需要根据导出的数据文件字段排列顺序进行定义  
配置文件中源数据文件和目的表需要根据情况定义

---

```
VERSION: 1.0.0.1
DATABASE: JCW
USER: gpadmin
HOST: 127.0.0.1
PORT: 5434
GLOAD:
  INPUT:
    - SOURCE:
      LOCAL_HOSTNAME:
        - 10.127.6.96
      PORT: 8088
      FILE:
        - /home/gpadmin/t_215001016.csv
    - COLUMNS:
      - sjsdccjmbbh: character varying(57)
      - yylx: character varying(7)
      - yhid: character varying(32)
      - zh: character varying(64)
      - qzid: character varying(128)
      - qzmc: character varying(128)
      - fszid: character varying(64)
      - fszzh: character varying(64)
      - fsznc: character varying(64)
      - jsxxnr: character varying(4000)
      - fssj: bigint
      - bddz: character varying(2)
      - ltjlid: character varying(32)
      - sczt: character varying(1)
      - scsj: bigint
      - fsxxlx: character varying(2)
      - zl_qzksjc: timestamp(6) without time zone
      - zl_dmql001: character varying(36)
      - zl_hcksjc: timestamp(6) without time zone
      - zl_is_problem: character varying(50)
      - zl_label_datasource: character varying(32)
      - zl_label_catalog: character varying(32)
      - zl_label_item: character varying(32)
      - zl_label_position: character varying(32)
      - zl_label_content: character varying(32)
      - zl_label_whitelist: character varying(32)
```

- zl\_score\_std: numeric(5,2)
- zl\_score: numeric(5,2)
- zl\_xzqh: character varying(6)
- zl\_ajbh: character varying(15)
- zl\_star: integer
- zl\_zhycxgsjc: timestamp(6) without time zone

- FORMAT: csv

- DELIMITER: ','

OUTPUT:

- TABLE: dzqz.qichengwentest
- MODE: INSERT

5. Master节点编写远程执行脚本test.sh, 用于远程执行gpload并清除数据文件

```
#!/bin/bash
```

```
#
```

```
gpload -f gpload1.yaml
```

```
rm -rf ./t_215001016.csv
```

6. etl服务器远端执行test.sh脚本完成数据加载

配置关键点