



```
-- 功能描述: DQC模板
-- 注意:
--     1. 不支持hive引擎,只支持spark引擎
--     2. 统计格式异常时不会判定null为异常
--     3. 僵尸表校验需要排除批次时间、时间戳和分区字段
--     4. 如果主键重复, 数据量校验、主键数据量递增、数值增长校验和僵尸表校验失效, null校验和格式校验不受影响
-- 校验点:
--     阻断: 主键重复校验、记录为空校验、记录格式错误校验、主键数据量递增校验 (选配)、数值增长校验 (选配)
--     提醒: 主键数据量波动校验、僵尸表校验
-- 校验表: credit_cust.dwd_cc_template
-- 输出表: credit_cust.dqc_cc_template
-- 主键   : pk1、pk2
with dwd_cc_template as
(
    select 'class-1' as pk1, 'tom'   as pk2, 70 as kpi, null      as phone, '20210101' as deadline, '${start|yyyyMMdd}' as dt union all
    select 'class-1' as pk1, 'tom'   as pk2, 80 as kpi, '1366666666' as phone, '20210000' as deadline, '${start|yyyyMMdd}' as dt union all
    select 'class-1' as pk1, 'tom'   as pk2, 90 as kpi, '13666666896' as phone, '20210000' as deadline, '${start-1d|yyyyMMdd}' as dt union all
    select 'class-2' as pk1, 'tom'   as pk2, 80 as kpi, '13666666896' as phone, '20210000' as deadline, '${start|yyyyMMdd}' as dt union all
    select 'class-2' as pk1, 'tom'   as pk2, 80 as kpi, '136666|6?66' as phone, '20210000' as deadline, '${start-1d|yyyyMMdd}' as dt union all
    select 'class-1' as pk1, 'jenny' as pk2, 60 as kpi, '13666666896' as phone, '90210101' as deadline, '${start|yyyyMMdd}' as dt union all
    select 'class-1' as pk1, 'jery'  as pk2, 50 as kpi, '136-689-886' as phone, '2021198' as deadline, '${start|yyyyMMdd}' as dt union all
    select 'class-1' as pk1, 'lisa'  as pk2, 40 as kpi, '13666666896' as phone, '2021/10/10' as deadline, '${start|yyyyMMdd}' as dt
),

dqc_cc_template as
(
    select concat_ws(',', pk1, pk2) as pk
        , sum(if(dt='${start|yyyyMMdd}',1,0)) as today_cnt
        , sum(if(dt='${start-1d|yyyyMMdd}',1,0)) as yesterday_cnt
        , max(if(dt='${start|yyyyMMdd}', kpi, null)) as today_kpi
        , max(if(dt='${start-1d|yyyyMMdd}', kpi, null)) as yesterday_kpi
        , sum(if(coalesce(phone,'')<>' ' and length(regexp_extract(phone,'(1[35678][0-9]{9})',0))<>11 and dt='${start|yyyyMMdd}',1,0)) as abnormal_phone_cnt
        , sum(if((unix_timestamp(deadline,'yyyyMMdd') is null or substr(deadline,1,1)>'2') and coalesce(trim(deadline))<>' ' and dt='${start|yyyyMMdd}', 1, 0)) as abnormal_date_cnt
        , if(count(if(dt='${start|yyyyMMdd}', concat_ws(',', pk1, pk2), null))>1, 1, 0) as duplicate_cnt
        , sum(if(dt='${start|yyyyMMdd}' and concat(pk1, pk2, phone, deadline) is null, 1, 0)) as null_cnt
        , if(max(if(dt='${start|yyyyMMdd}', md5(substr(concat_ws(',',t.*),1,length(concat_ws(',',t.*))-8)), null))
            =max(if(dt='${start-1d|yyyyMMdd}', md5(substr(concat_ws(',',t.*),1,length(concat_ws(',',t.*))-8)), null)), 1, 0) as jiangshi_cnt
    from dwd_cc_template t
    where dt in('${start|yyyyMMdd}','${start-1d|yyyyMMdd}')
    group by concat_ws(',', pk1, pk2)
)

select from_unixtime(unix_timestamp()) as insertdt
    , sum(today_cnt) as today_cnt -- 今日记录条数 (主键不重复时等于主键数量)
    , sum(yesterday_cnt) as yesterday_cnt -- 昨日记录条数 (主键不重复时等于主键数量)
    , sum(abnormal_phone_cnt) as abnormal_phone_cnt -- 今日手机号格式异常记录数量 (主键不重复时对应主键)
    , sum(abnormal_date_cnt) as abnormal_date_cnt -- 今日日期格式异常记录数量 (主键不重复时对应主键)
    , sum(duplicate_cnt) as duplicate_cnt -- 主键是重复的主键数
    , sum(null_cnt) as null_cnt -- 今日关键字段为null的记录数量 (主键不重复时对应主键)
    , if(sum(duplicate_cnt)>0, 0, 1) as chk_pk -- 主键重复校验
    , if(sum(duplicate_cnt)=0 and (sum(today_cnt)<sum(yesterday_cnt)*0.6 or sum(today_cnt)>sum(yesterday_cnt)*1.4), 0, 1) as chk_cnt_range -- 主键数据量波动校验
    , if(sum(duplicate_cnt)=0 and sum(today_cnt)<sum(yesterday_cnt), 0, 1) as chk_cnt -- 主键数据量递增校验
    , if(sum(today_cnt)<>0 and sum(null_cnt)>0, 0,1) as chk_null -- 记录为空校验
    , if(sum(abnormal_phone_cnt)+sum(abnormal_date_cnt)>0, 0, 1) as chk_format -- 记录格式错误校验
    , if(sum(duplicate_cnt)=0 and sum(today_cnt)<>0 and sum(jiangshi_cnt)=sum(today_cnt), 0, 1) as chk_jiangshi -- 僵尸表校验
    , if(sum(if(today_kpi<yesterday_kpi and yesterday_kpi is not null, 1, 0))>0, 0, 1) as chk_value_rise -- 数值增长校验
from dqc_cc_template t
```