拟定6月1日正式开始启动所有服务器的动态数据采集

|  |  |  |
| --- | --- | --- |
| 表名称 | 更新规则 | 备注 |
| 1.主播根目录 | 在5月30日建立好各个服务器里的主播跟目录里的所有主播url | 只是最初建立，以后不更新 |
| 1.1主播信息 | 从6月1日起每天都更新 | 早8点重启 |
| 1.2表现趋势 | 在5月31日抓取相应近90天的粉丝数、获赞数、作品数的趋势数据（应该一天可以抓完吧？） | 以后不更新 |
| 1.3作品列表 | 在5月31日抓全从1月31日到5月30日（累计4个月的）作品url的相关数据（应该一天可以抓完吧？）  从6月1日期开始每日更新，每天抓取昨天的作品url等数据（如6月1日抓取5月31日的） | 早8点重启 |
| 1.3.1作品列表\_评论 | 从6月1日期开始每日更新，每天抓取是4个月前记录的作品url的数据（如6月1日抓取的是1月31日的；6月2日抓取的是2月1日的）2月没30号这个不会出错吧？ | 早8点重启 |
| 1.3.2作品列表\_趋势表现  1.3.3作品列表\_参与者画像  1.3.4作品列表\_关联商品 | 从6月1日期开始每日更新，每天抓取是4个月前记录的作品url的数据（如6月1日抓取的是1月31日的；6月2日抓取的是2月1日的） | 早8点重启 |
| 1.4到1.4.10 | 从6月1日期开始每日更新，每天抓取昨天的数据（如6月1日抓取5月31日的） | 早8点重启 |

补充说明：

* 由于主播每天可能播好几场，为了方便区分，请将表1.4到1.4.10里的amount\_zb替换为livestraming\_time
* 表1.4.9实时观众停留时长具体的时长计算规则如下：

Timepoint记录的是实时流量表中的时间点（如16:57:33，17:00:50，17:02:34）

shichang记录的是间隔时长，对应计算规则是（直播开始时间到16:57:33，16:57:33到17:00:50，17:00:50到17:02:34）

* 表1.4到1.4.10虽然每日更新但没法抓取之前的历史数据，能否单用1台服务器依次抓取12个类别直播从1月31日到5月30日历时4个月的历史数据（就是将12个类别的根目录集成到一起，依次抓4个月的历史数据，抓完就停止）
* 具体服务器的分配情况，可能要看第一台的实际能力才能确定