

抽取日志

抽取日志跟踪

日志类型

目前抽取过程的日志分为两种：

- 1、抽取过程的错误日志，这种是在抽取异常的时候会记录在error文件，在数据集抽取日志中能查看。
- 2、抽取过程中的DEBUG日志，这种是为了便于跟踪抽取过程直接输出到控制台。

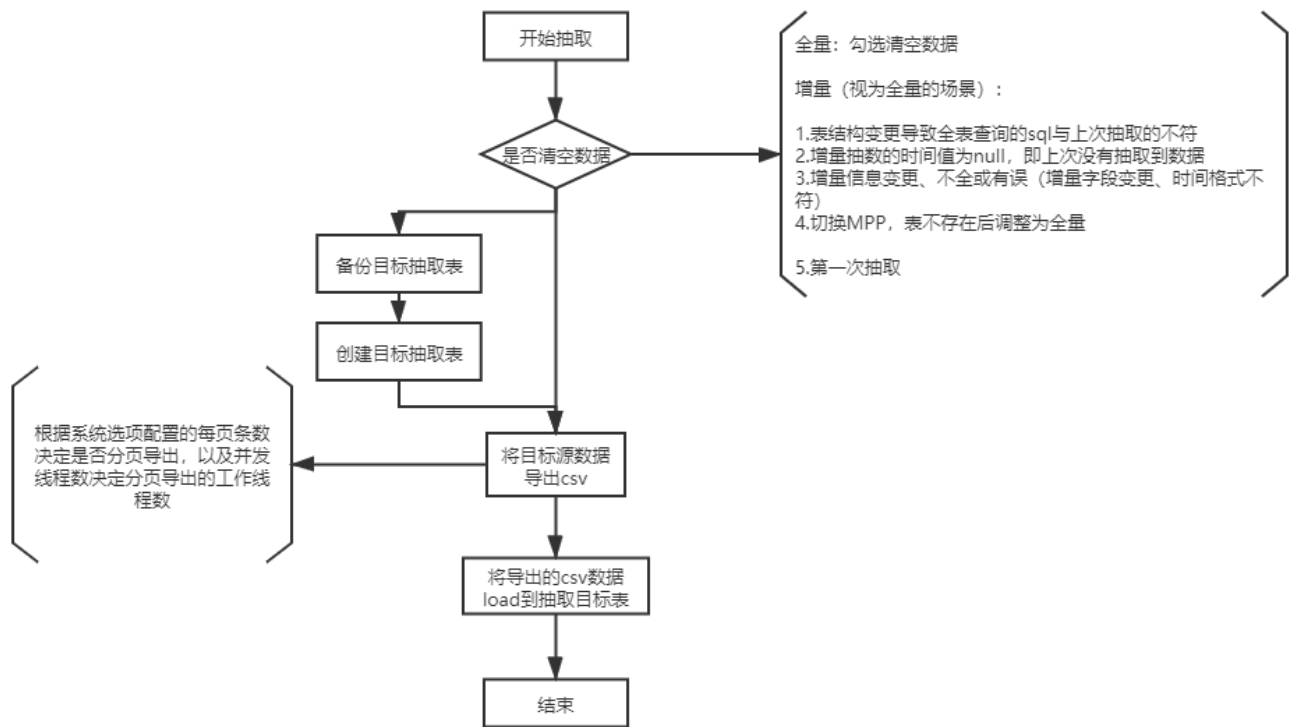
- 抽取日志跟踪
 - 日志类型
 - 错误日志
 - DEBUG日志
- 抽取过程
- 示例
- 注意事项

错误日志

错误日志我们可直接在抽取监控面板查看或者右键 **数据集 > 调度管理 > 抽取日志查看**。

DEBUG日志

抽取过程的流程图，如下：



DEBUG日志，我们有两种方式实时监控：

- 1、tail -f smartbi.log
- 2、系统监控-日志，这里可以在浏览器实时监控

抽取过程







整理抽取整个过程的日志，分为以下几个步骤：

以控制台日志输出以及异步写日志表的方式实时监控抽取进度

- 1、获取总行数：记录耗时

控制台：- **【获取总行数】**，耗时：2353ms，本次抽取已耗时2807ms







日志表：

	 c_endtime	 c_type	 c_lines	 c_state	 c_log	
17	2020-05-18 20:00:31	FULL	830	DONE	数据抽取成功	[N]
18	2020-05-18 20:03:19	FULL	830	DONE	数据抽取成功	[N]
19	2020-05-18 20:07:28	FULL	1100000	DONE	数据抽取成功	[N]
20	2020-05-18 20:10:21	FULL	1100000	DONE	数据抽取成功	[N]
21	2020-05-18 20:42:24	FULL	1100000	DONE	数据抽取成功	[N]
22	2020-05-18 20:57:16	FULL	0	RUNNING	最近一次在执行【获取总行数】，耗时：2353ms，本次抽取已耗时2807ms	[N]

2、单线程导出的取数：记录取数耗时

控制台：- 【单线程导出取数】，耗时：20ms，本次抽取已耗时503ms



日志表：

	 c_type	 c_lines	 c_state	 c_log	 c_csvpath	 c_faillogp
19	FULL	1100000	DONE	数据抽取成功	[NULL]	[NULL]
20	FULL	1100000	DONE	数据抽取成功	[NULL]	[NULL]
21	FULL	1100000	DONE	数据抽取成功	[NULL]	[NULL]
22	FULL	1100000	DONE	数据抽取成功	[NULL]	[NULL]
23	FULL	1100000	DONE	数据抽取成功	[NULL]	[NULL]
24	FULL	0	RUNNING	最近一次在执行【单线程导出取数】，耗时：20ms，本次抽取已耗时503ms	[NULL]	[NULL]

3、导出csv：记录每个csv文件的导出条数、导出耗时

控制台：- 【多线程导出到CSV】，导出条数：100000，总条数：1100000，耗时：4341ms，本次抽取已耗时76038ms



日志表：

	 c_lines	 c_state	 c_log	
17	830	DONE	数据抽取成功	
18	830	DONE	数据抽取成功	
19	1100000	DONE	数据抽取成功	
20	1100000	DONE	数据抽取成功	
21	1100000	DONE	数据抽取成功	
22	100000	RUNNING	最近一次在执行【多线程导出到CSV】，导出条数：1000000，总条数：1100000，耗时：26992ms，本次抽取已耗时98682ms	

4、导入目标表：记录每个csv文件的导入条数、导入进度百分比

控制台：- 【CSV导入到目标表】，导入100000条数据，导入进度：100%，本次抽取已耗时78851ms

日志表：

	 c_type	 c_lines	 c_state	 c_log	
17	FULL	830	DONE	数据抽取成功	
18	FULL	830	DONE	数据抽取成功	
19	FULL	1100000	DONE	数据抽取成功	
20	FULL	1100000	DONE	数据抽取成功	
21	FULL	1100000	DONE	数据抽取成功	
22	FULL	1100000	RUNNING	最近一次在执行【CSV导入到目标表】，导入1000000条数据，导入进度：100.00%，本次抽取已耗时116046ms	

示例

以下是一条完整的抽取过程

1、获取总行数：

耗时：2293ms，本次抽取已耗时2801ms

2、多线程导出到CSV：

导出条数：100000，总条数：1100000，耗时：6277ms，本次抽取已耗时11116ms

3、CSV导入到目标表：

导入100000条数据，导入进度：9.09%，本次抽取已耗时16241ms

4、多线程导出到CSV：

导出条数：1000000，总条数：1100000，耗时：41649ms，本次抽取已耗时46486ms

5、CSV导入到目标表：

导入1000000条数据，导入进度：100.00%，本次抽取已耗时69163ms

注意事项

Vertica由于不支持多文件分批处理，暂不支持。