克拉玛依数据对账

1. 克拉玛依数据对账计划做如下四张图表：
2. 汇总

【税号】 【开始时间】【结束时间】 查询按钮

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 税号 | 盘号 | 税盘数据量 | 中间库数据量 | 大数据平台数据量 | 起始时间 | 终止时间 |
| tax1 | disk1 | 50 | 50 | 50 | 2019/08/04 | 2019/09/03 |
| tax2 | disk2 | 60 | 59 | 58 | 2019/08/04 | 2019/09/03 |

表 1

区间时间内数据量汇总图，默认为近一个月内数据。按天维度统计汇总

提供两个筛选条件：1. 提供单税号查询，2. 支持更改起始时间查询更长时间的汇总数据。

1. 饼图

根据图一中的税盘数据量，中间库数据量，大数据平台数据量，可以得出两个比例：中间库数据量/税盘数据量，和 大数据平台数据量/中间库数据量，

根据这两个比例，可以得到下面的两个饼图：

图 1

饼图（样例数据）

根据图图 1可以直观的看出大概的数据抽取情况，以及大部分数据问题出在中间件挖油，还是助手上传过程中。

点击饼图的任一部分，得到筛选过的相对应比例的那部分数据结果，格式同图一一致，点击单个税号，显示出此税号默认一个月内每天的数据情况图(图三)。

1. 单税号每天数据情况图

【开始时间】 【结束时间】 查询按钮

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 日期 | 税号 | 盘号 | 税盘数据量 | 中间库数据量 | 大数据平台数据量 | 详情按钮 |
| 2019/09/03 | tax1 | disk1 | 10 | 8 | 7 | [详情](xxx) |
| 2019/09/02 | tax1 | disk1 | 10 | 8 | 7 | 详情 |
| … | … | … | … | … | … | … |
| 2019/08/04 | tax2 | disk2 | 8 | 8 | 8 | 详情 |

表2

图三，单税号每天数据量情况，默认为前一天所有税号的数据情况；

提供税号输入框和时间选择框，可以查询指定税号前一天的数据情况，也可以查询单个税号一个时间区间内的所有天的数据情况列表。

最后一列详情为超链接，点击显示单税号的挖油详情(图四)。

1. 单税号挖油详情图

【税号】【版本号】【查询按钮】

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 日期 | 税号 | 盘号 | 登录时间 | 登录时长 | 版本 | MAC | 税盘数据总量 | 中间库数据总量 | 大数据平台数据总量 |
| 2019/09/03 | tax1 | disk1 | 08:10 | 15 | 1.0.0.12 | xxx | 769 | 769 | 760 |
| 2019/09/03 | tax1 | disk1 | 09:20 | 24 | 1.0.0.12 | xxx | 778 | 778 |
| 2019/09/03 | tax2 | disk2 | 10:20 | 5 | 1.0.0.11 | xxs | 87 | 78 | 78 |

表 3

表3，税号挖油详情图

可以直接由图三点击详情链入，也有单独页面查询；

单独页面查询时，默认显示前一天所有税号的详情数据，提供指定税号，和指定时间区间内的查询。

PS：除饼图外，上述几张图单图默认数据量在600万左右。

1. 数据补全

短期先人肉观察，然后人肉触发。 客户端实现a:全量补偿(重抽)，b 增量补偿

短期内，人肉对问题数据进行分析查询原因，暂定一个月内，得出相应的问题判定条件，及相应的数据补全策略，如重置挖油时间，重新挖油，或者重新挖取某天的油；

规则敲定后，由大数据根据相应的规则，每天或者每周推送问题数据给客户端分析处理。

三、 smartbi和嵌入运营管理后台优劣对比

|  |  |  |
| --- | --- | --- |
| 方案 | 优点 | 缺点 |
| Smartbi | 1. BI工具，饼图易做 | 1. 多图联动难做； 2. mysql查询时，自动嵌套成子查询，增加查询复杂度，占用更多资源； 3. 对hive的支持效率待验证； 4. 大数据量难以承载，无法分页，只能全部显示出来，滚动查询，但是滚动一屏就会出现‘拼命加载中’字样，长时间会超时； 5. 无法做超链跳转，点击下钻到其他图表； 6. BI工具，主要用于数据结果的趋势图形化展示，不建议用来做大数据量的数据对账查询工具。 |
| 嵌入运营管理后台 | 1. 支持hive； 2. 查询sql语句自定义； 3. 支持自定义分页； 4. 自定义点击事件，超链跳转。 | 1. 做饼图时，需借助Echarts等第三方开源代码； |

为什么选择自己开发页面：

1. 排查问题的时候更多的时候想做到多表之间的联动， 如饼状图跳转到了细节表(表2)。 细节表的详情跳转到详情表（表3）
2. 对表格因为客户端一天有多次登录的需求，所以有合并单元格的工作，如(表3)
3. smartbi的表格没找到分页的方式， 对于大量的数据显示性能有问题。且查询条件控制不易