****

**平台开发规范V1.0**

*修订历史*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 版本号 | 修订日期 | 修订人 | 审核日期 | 审核人 | 修改说明 |
| V1.0 | 2016/11/2 |  | 2016/11/2 |  | 修改 |
| V1.1 | 2016/11/23 |  | 2016/11/23 |  | 新增 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**为了更好的服务平台预上线项目，加快项目的上线进度，提高工作效率，避免项目的多次修改与测试，对开发的程序进行规范，规范如下：**

**一、规范信息**

**1 、程序架构要求**

|  |  |
| --- | --- |
| 场景描述 | 生产区平台程序上线，对程序本身架构要遵循规范。 |
| 规范内容 | 程序建议采用mr的架构进行开发。 |

**2 、程序评估要求**

|  |  |
| --- | --- |
| 场景描述 | 程序上线前应提供平台所需基础信息，为资源评估所用。 |
| 规范内容 | 应提供程序运行时所需的hadoop组件版本以及其它环境要求，以评估现有环境是否符合要求。 |

**3、dpi数据加载要求**

|  |  |
| --- | --- |
| 场景描述 | 如果程序需要加载DPI数据要遵守以下规范。 |
| 规范内容 | 建议程序如需访问dpi数据时，按照小时来进行加载，避免因一次性加载数据量过大而无法加载问题出现。 |

**4、数据加载内容要求**

|  |  |
| --- | --- |
| 场景描述 | 如果程序加载数据省份发生量级转变要遵守以下规范。 |
| 规范内容 | 建议程序制定省份优先级别，可以在程序稳定逐步添加省份，避免因一次性加载数据量过大而无法加载问题出现。 |

**5、文件识别要求**

|  |  |
| --- | --- |
| 场景描述 | 程序本身会对tmp下文件进行操作要遵守以下规范。 |
| 规范内容 | 开发程序要能识别tmp文件并能正确读取。 |

**6、结果数据读取要求**

|  |  |
| --- | --- |
| 场景描述 | 程序处理出结果数据后，后续接口需要读取要遵守以下规范。 |
| 规范内容 | 程序在hadoop部分处理完后，如果后续接口程序需要用到此部分数据，应直接从hdfs上读取，不允许把hdfs上数据导到本地系统再进行读取。 |

**7、日志存放要求**

|  |  |
| --- | --- |
| 场景描述 | 如果程序有日志输出要遵守以下规范。 |
| 规范内容 | 程序中不能写死日志存放的目录，日志存放的目录可以调整。 |

**8、程序结果数据清理要求**

|  |  |
| --- | --- |
| 场景描述 | 如果程序会输出结果数据，应对此类数据进行及时的清要遵守以下规范。 |
| 规范内容 | 程序应有删除机制，在程序运行完后及时清理数据，如果需要对数据保留，应制定保留期，保留期存储的数据不超过销售合同的存储，不能无限期保留。 |

**9、数据访问要求**

|  |  |
| --- | --- |
| 场景描述 | 如果程序有访问数据的需求，要遵守以下规范。 |
| 规范内容 | 提供需要访问的数据以及在本地和hdfs上需要创建的目录。 |

**10、程序及配置文件的大小要求**

|  |  |
| --- | --- |
| 场景描述 | 如果程序、配置文件、脚本等内容过多，要遵守以下规范。 |
| 规范内容 | 程序以及相应的配置文件，脚本不应过大，建议在1G以内。 |

**11、程序及配置文件的书写要求**

|  |  |
| --- | --- |
| 场景描述 | 如果程序和配置文件是一体的，要遵守以下规范。 |
| 规范内容 | 配置文件应和程序分开，不能写死在程序中，可以避免后续由于业务的更改而需修改程序。 |

**12、程序结果数据大小要求**

|  |  |
| --- | --- |
| 场景描述 | 如果程序有输出结果，要遵守以下规范。 |
| 规范内容 | 程序运行后的结果文件不能是小文件，文件的大小需大于128M。 |

**13、程序结果数据大小要求**

|  |  |
| --- | --- |
| 场景描述 | 如果程序有输出结果，要遵守以下规范。 |
| 规范内容 | 程序运行后的结果文件不能是小文件，文件的大小需大于128M。 |

**14、入库程序架构要求**

|  |  |
| --- | --- |
| 场景描述 | 如果程序有入库需求，要遵守以下规范。 |
| 规范内容 | 入库程序需是分布式入库，即是程序提交后任务运行在集群上，而不是在客户端。 |

**15、需平台配合排错的流程要求**

|  |  |
| --- | --- |
| 场景描述 | 如果程序出现问题，需要平台组配合解决，要遵守以下规范。 |
| 规范内容 | 运维只保证集群的稳定，如开发测试遇到报错需要运维配合排查时，应由接口人发需求工单。 |

**16、程序使用hive组件的要求**

|  |  |
| --- | --- |
| 场景描述 | 如果程序需要使用到hive，要遵守以下规范。 |
| 规范内容 | 如程序中需要使用hive,目前平台只提供hive client的使用，不开放hive server的使用。 |