# ETL功能需求

1. 读取文件组件，适用于一般顺序文件（定长或不定长），可识别文本文件、excel、xml等文件。单一或多文件输入，单文件输出，能够识别各种自定义分隔符，文件内一条记录的结束、配置好后能够预览数据，读多个文件时，采用并行方式，单个文件则顺序方式，缺省情况下，一个节点读取一个完整的文件。可以定义多节点同时读取文件已提升性能，写多个文件时，采用并行方式，单个文件则顺序方式，缺省情况下，一个节点写一个完整的文件，单个节点可以写多个文件。支持动态字段的定义。
2. 拷贝组件，可以有一个输入，多个输出。它可以在输出时改变字段的顺序，但是不能改变字段类型。
3. 筛选组件，只有一个输入，可以有多个输出。根据不同的筛选条件，可以将数据输出到不同的输出连线上。
4. 合并组件，将多个字段相同的数据文件合并为一个单独的文件输出。可以按照指定的key值排序。
5. 转换组件，有一个输入，多个输出，可以将字段进行转换，也可以通过条件来指定数据输出到那个输出连线上。在开发过程中可以使用拖拽。并且可以在每个输出连线上写条件控制数据的转换输出，每个字段都可以映射到输出端。
6. 排序组件，只能有一个输入及一个输出，按照指定的Key值进行排列。可以选择升序还是降序，是否去除重复的数据等等。
7. 数据关联组件，把数据读入内存执行查询操作，将匹配的字段输出，或者在在符合条件的记录中修改或加入新的字段。
8. 连接组件，将多个表连接后输出。内连接：相同key的数据输出，不同的删除，左连接：保留所有的left数据和其他输入数据key匹配的列，并删除其他输入数据的key，右连接：保留所有的right数据和其他输入数据key匹配的列，并删除其他输入数据的key，全连接：输出两个连接中所有的。
9. 合并组件，将key值相同的记录合并。
10. 修改组件，只能有一个输入及一个输出，它可以修改表结构：删除，保留一个字段；给字段改名；修改字段的数据类型。
11. 写文件组件，将数据写入文件。
12. 连接主流数据库组件。如：oracle ，DM，DB2，MYSQL等。
13. 分组计算组件，将输入的数据分组，计算各组数据的总和或者按组进行其他的操作，最后将结果数据输出到其他的组件。
14. 去重复数据组件。输入根据关键字分好类的有序数据，去除所有记录中关键字重复的记录。
15. 压缩与解压组件。
16. 比较组件，按字段比较两个文件，找出不同的记录。
17. 分割组件，将文件按照一定的条件（一般为字段的值）分割成多个子文件。具体是将输入的每一条记录按照各自符合的条件（关键字的值）分配到不同的输出。
18. 调用存储过程，shell命令组件。
19. ETL调度功能如下：

用户管理：新建、修改、删除用户组和用户以及用户组和用户的权限配置；

应用管理：新建、修改、删除应用主题与应用；

集群管理：配置集群节点属性，包括集群基本信息、参数配置、状态设置、日切事件管理；

ETL域管理：新建及配置ETL域、ETL Server、ETL Client；

部署管理：提供环境间参数的导入、导出；

全局变量管理：新建，修改，删除全局变量；

系统节假日管理：可定制系统级节假日与工作日；

日志查询：查看登录系统用户的相关信息与操作；

作业资源类型管理：新建、修改、删除作业资源类型并为调动作业配置资源类型。

监控管理：

数据监控：数据批次、数据实例运行情况

日常任务监控：监控日常调度任务及其批次、作业流，作业的运行情况，包括运行统计信息和图形化展示。