

# ETL工具比较 (Informatica ,SSIS,Kettle ) - - 博客频道 - CSDN

## Informatica VS SSIS

从易用性来说, Informatica有四个开发管理组件, 初学者可能不知道如何下手, 所以关键先要了解PowerCenter的开发流程。SSIS有BIDS和SSMS, 主要的开发工作在BIDS上完成, 但是BIDS是基于 Visual Studio的, 如果是没有做过.NET开发的人员, VS的复杂的开发环境可能让他们不知所措, 但是对于熟悉VS的开发人员来说, 通过VS, 他们可以使用更多的资源, 比如源管理器, VSTS等。从界面的人性化上看, SSIS无疑更好, 这是Microsoft的一贯特色, PowerCenter有些地方连右键菜单都没有。

SSIS的数据流如同PowerCenter的Mapping, SSIS的控制流如同PowerCenter的Task, SSIS的Package如同PowerCenter的WorkFlow。PowerCenter的Mapping和WorkFlow分开开发, 在一个大型开发团队中, 开发人员开发Mapping, 而有管理人员将多个Mapping在WorkFlow中统一调度。但如果是只有几个人的开发团队中, 这种模式的效率就不如SSIS了, 因为小型团队中每个人都独立工作, 开发人员需要开发好Mapping后, 再在WorkFlow建立Session, 需要再一次为Session配置连接和参数。相反, SSIS不适合多人开发, 因为不像PowerCenter将Mapping和WorkFlow分开开发, 所以当一个人开发Package时, 其他人就插不上手。

SSIS的运行监控是非常棒的设计, 可以看到数据流的执行情况。如果数据流出问题, 可以马上知道在那个转换组件上发生了问题。而PowerCenter的Monitor只能看到源和目标的情况, 一旦出状况, 就只能看日志了, 但在日志大量的文本里, 很多时候是找不出问题的所在。除了运行时监控外, SSIS在开发时可以动态的Validate, 而且它的Validate也严格得多。只要发现字段类型对应不上就会报错。PowerCenter中Port的对应比较自由, 比如一个Output的Port可以随意对应到多个Input的Port, 而SSIS中一定要用多播组件才行。所以在数据质量方面, SSIS比PowerCenter更有保障, PowerCenter最郁闷的是, 报错后却找不到出错的地方。

就组件的功能上, SSIS比PowerCenter要丰富, 像SSIS控制流的循环组件, 数据流中的模糊查找、词频计算、数据挖掘评分等组件都是Informatica没有的。特别是有很多针对SQL Server组件。如果使用非SQL Server数据库的话, SSIS功能上要打个七折。

SSIS最致命的问题是数据库连接的兼容性上, 数据流中很多组件只认OLE DB连接, 而且只有OLE DB数据源组件才可以加参数, 这显然是Microsoft有意为之。加上SSIS的Validate比较严格, 所以连接到那些对OLE DB兼容不好的数据库就会报错, 像我之前连接Oracle 9i就碰到这样的问题。PowerCenter支持更多的数据源, 此外还可以连接SAP, PeopleSoft等商业软件, 在大型企业中, 这些功能是非常方便的。

## Informatica VS Kettle

- 1、 Informatica属于商业软件, 而Kettle是开源软件; Informatica的收费方式是软件一次性购买, 但技术支持按项目收费。具体怎么界定项目不清楚。
- 2、 风险和成本方面自然就不用说了, 从某些方面来说的话, 风险总是与成本成反比的;
- 3、 易用性方面, Informatica和Kettle都有GUI图形界面, 操作步骤都比较简单易用; 主要取决于开发人员的对工具的熟悉情况, 没有太大的可比性;
- 4、 Kettle是基于Java开发的ETL工具, 在使用过程中需要借助JVM, 在数据抽取速度以及大数据处理能力方面远远不如Informatica; (Kettle在异构数据库的处理能力比同构数据库的处理能力要弱很多, 特别是跨服务器之间就更弱了)
- 5、 Informatica将客户端和服务端进行分离, 而Kettle则没有但需要基于JVM; 因此Kettle的部署比较简易;
- 6、 稳定性方面, Informatica也有较大的优势;
- 7、 Informatica将元数据及一些配置信息存储于配置的数据库, 也就是说Informatica需要借助其他数据库才能运行;
- 8、 Kettle的数据转换过程相对比较灵活, 可以是手工编写的SQL语句、Java代码、正则表达式等; Informatica则由Data Quality组件来保证数据的质量, 当然也支持SQL语句, 两者没有明显的差异;
- 9、 Informatica和Kettle均有监控日志功能, 这也是ETL工具必须具备的, Informatica的监控日志相对要详细一些, 但在实际应用中, 大多数情况下不需要这么详细;
- 10、 数据源方面, 二者没有明显区别, Kettle支持广泛的数据库、数据文件等, 同时支持扩展; Informatica广泛支持各类结构化、非结构化的数据源;

