

SQL2005中的SSIS与Oracle数据库的迁移性能 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/207/2021_2022_SQL2005_E4_B8_AD_c102_207487.htm 一直担心sql server 2005的ssis性能问题，一位仁兄对其进行了测试，值得借鉴。项目中存在一部分数据迁移的工作，说白了就是从老的系统中将数据倒换的新的系统模型中，老系统的数据来源比较复杂多样，新的自然是Oracle9.2。本来这也就是一次性工作，用SQL自然是最快的方式，不论是开发还是数据传输的速度。可是甲方偏偏要看到界面，希望这是一个成型的工具，没办法，甲方就是上帝。公司原来也有一个迁移工具，可是只能适用于表对表的倒换，复杂一些无能为力，而且数据还巨慢，用过的人都是对它无语。从新开发，不说花费和效果，光是时间也不行。没办法，只好看看现在流行的ETL的工具。市场前列毋庸置疑，肯定是Informatica 和 DataStage. Informatica没有，只好看看DataStage是否能适应现在的功能要求。不想，虽然是图形界面，可使用起来一点也不容易，而且安装后，Windows下居然不能脱离域环境，而且不是Server版本的Windows还不能运行Paralle Job。郁闷无比。试了两天后，暂时放下。Microsoft的易用性比功能强大更吸引我。试试SQL Server 2005中的SSIS，号称企业级的ETL。一用之后呢，没想还真有点喜欢上了它，从介绍的和界面上看一点也不比DataStage的功能少，性能，哈，下面就是我要说得了。ETL工具最慢的部分都是L这一部分，按照一般的说法能占到总体时间的五分之四，所以这是关键。测试也不算复杂，就是同样的数据抽取、转化、然后加载用不同的驱动分别跑一遍，目的库已经

确定是Oracle，所以也没有太大的余地了。在SSIS中，有两个驱动可以连接Oracle数据库，一个是Microsoft OLEDB Provider for Oracle，另外一个Oracle Provider for OLEDB不测不知道，还真长了不少见识。同一机器，同一数据源，同一结果，两者间还真有不少区别。首先是速度（连续三次）：
Microsoft OLEDB Provider for Oracle 1分37 1分32 1分30 Oracle Provider for OLEDB 1分10 1分07 1分02 在速度上 Oracle Provider for OLEDB 基本符合 1分3万条左右，而Microsoft OLEDB Provider for Oracle 1分钟只有2万条左右。照这样看，答案似乎也就出来了，Oracle Provider for OLEDB也就成了不二选择。且慢，我还没有说明为什么选择25万条记录而不是别的数量的数据呢。这就不得不说说内存的使用：未启动数据迁移时即停留在VS.Net设计界面时，内存已使用了790M左右，而我机器的物理内存也就896M。运行开始后，25万条记录下Microsoft OLEDB Provider for Oracle 平均在1G左右，而Oracle Provider for OLEDB乖乖得不得了，铁定在1.25G以上，一次还在1.3G。更离谱的是，原数据表中共有近100万条记录，Microsoft OLEDB Provider for Oracle在内存峰值1.5G左右可以顺利完成，而Oracle Provider for OLEDB在内存使用一旦突破1.3G往上一些，就开始不停提示内存不足，不在安心的迁移数据了，或者干脆显示为红色，报一些莫名的错误。这就让人两难了，一个速度快了那么50%，可确是一个内存消耗大户，有没有止境，我这破机器也无从得知。另外一个速度慢，可却节俭持家，穷人也照顾到了，哈。感觉好这有点像Oracle和MS的企业风格，一个走高端，为了需要的指标可以不计成本，穷人靠边；另一个呢，还不错，虽然也越来越

来不鸟没钱的人，可还做得不太显眼。最后了，同样的数据源（Microsoft OLEDB Provider for Oracle驱动），将目的库换成SQL Server 2005,驱动为SQL Native Client,同样的数据数据转换，98.9万条记录中11.1万条入库，靠1分12完事，打开FastLoad，58秒搞定。而且都只是第一次运行，相信如果多运行几次后，结果应该更好。别说，自家孩子真就不一样，别人的家的没法比。100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com