Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Study of ETL Tools: Talend Open Studio and SQL Server Integration Services (SSIS)

Premkumar Tejnani[1]
[1]P. G Student, Department of MCA, VES Institute of Technology, Chembur, Mumbai, Maharashtra, India

Abstract: ***ETL basically stands for extraction, transformation and loading. It is a process which includes extracting the data from source, transformation of that data and loading that data into the target tables or files. This paper basically studies the two ETL tools which are talend open studio and the SQL server integration services (ssis) and differences are discussed between the two tools.***
keywords: Data warehouse, TALEND, ETL, OLAP, SSIS, SQL.

I. INTRODUCTION

Data warehouse sometimes referred as Enterprise data warehouse is a system which is used for reporting and analysis of data. One of the most common type of data warehouse is ETL based data warehouse which uses various types of layers such as staging, data integration and access layers. Using data warehouse we can have the data in multidimensional view and we can also make use of OLAP tools for analyzing data in the multidimensional space. The concept of data warehousing has served the need to have access to a structured data in an easy manner. The term ”Data Warehouse” was first coined by BillInmon in 1990 [1]. The data warehouse primarily contains the historic data of the organization which stored so that it can be later used for data analysis.

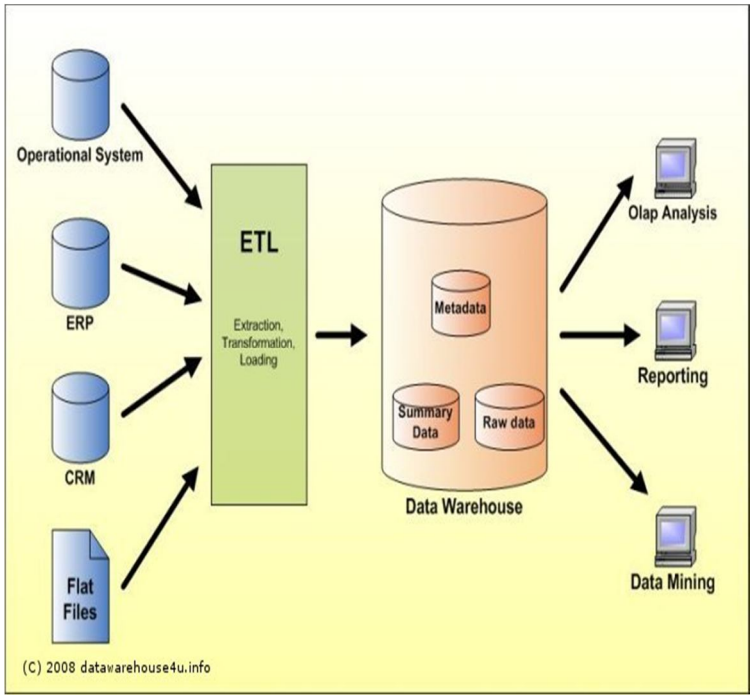

Fig.1 Overview of data warehouse.

II. ETL PROCESS

In today’s world, the major decisions which are taken in the organization depend on the data which is store in various types of system in the organization. For converting the data into useful information, that data needs to be accumulated from various is usually preferred for loading small amount of data. And bulk loading is preferred for loading large amount of data.

Volume 5 Issue VI, June 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

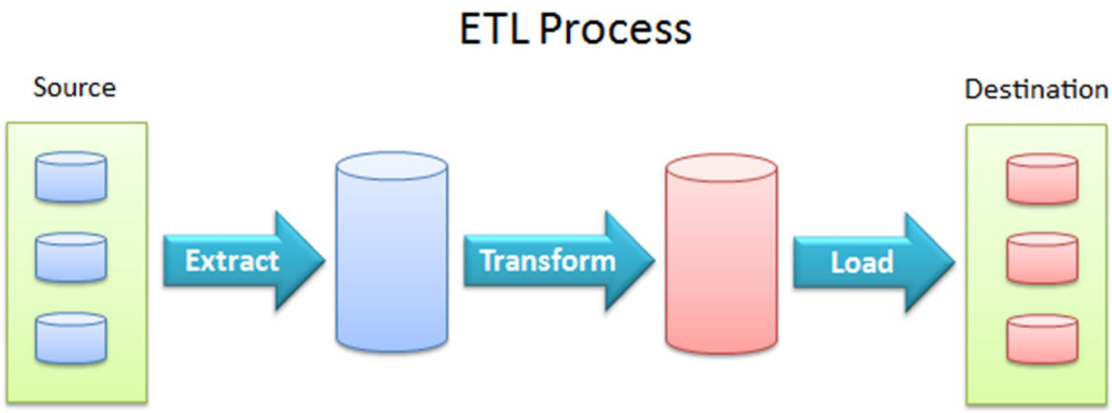

Fig. 2 ETL Process

III. TALEND OPEN STUDIO

Talend is the leading open source integration software provider to data-driven enterprises [2]. Talend Open Studio is an open source ETL tool. IT is a data integration software released in 2006.It is used for two purposes. First is integrating the operational systems and second is as an ETL tool for data warehousing. It is designed to meet the data integration needs of all types of organizations. It is one of the most powerful open source data integration tool available in the market today

A. Benefits

1) It is an open source ETL tool.
2) Give the ability to access and extend source code to users needs.
3) One of the most powerful open source data integration tool available.
4) Backed by Talend's ongoing and extensive Research and Development.
5) It is frequently updated and enhanced in many ways.
6) Java code can be generated from the developed packages
7) It is free to download.

Its GUI gives access to a meta data repository and to a graphical designer. The meta data repository contains the definitions and configuration for each job - but not the actual data being transformed or moved. All of the components of Talend Open Studio for Data Integration use the information in the meta data repository.

IV. SQL SERVER INTEGRATION SERVICES (SSIS)

SQL Server Integration Services (SSIS) is a component of the Microsoft SQL Server database software that can be used to perform a broad range of data migration tasks [3]

SSIS is one of the Business Intelligence tools. It was developed by Microsoft corporation. Microsoft Integration Services is a platform for building enterprise-level data integration and data transformations solutions [4].It was designed to ease and automate the process of ETL. The advantage that SSIS has over other ETL applications is that it is not just limited to data warehousing. For example and SSIS package can be created to automate SQL server maintenance plan. It can be used by anyone from commercial consultants to individual researchers. This product has been designed to provide better approach towards data migration, manipulation and transformation. user can easily define how the process should flow and perform some task on different interval

A. Advantages of SSIS

1) can handle data from heterogeneous data sources at a same package.
2) SIS consumes data which are difficult like FTP, HTTP, MSMQ, and Analysis services etc.
3) Remove network as a bottleneck for insertion of data by SSIS into SQL.
4) Easy to maintain and package configuration.

5) Tightly integrated with Microsoft Visual Studio and SQL Server.
6) Data can be loaded in parallel to many varied destinations.
7) Remove network as a bottleneck for insertion of data by SSIS into SQL.

V. DISCUSSION/COMPARISON

SSIS is known for various capability. It mainly focuses on the delivery of large amount of data. According to today's market needs, this functionality is very much required from an ETL tool. It is very fast and easy to implement the SSIS. A huge advantage of using SSIS over other ETL tools is its ability to integrate it with Microsoft Office. On the other hand Talend is a new Company. Talend delivers open source products. This lets the user modify the source and create a custom product which serves there needs. Talend products are purchased by customers worldwide. The integration capabilities of Talend with other platforms are much efficient than the SSIS. And SSIS when compared with other ETL, is very cost effective and it also has various data management capability. Talend is suitable for small implementations where as SSIS is suitable for large implementation. Talend is new to market where as Microsoft is well established company which has experts. Another reason for success of SSIS is the popularity of Microsoft. So most of the customers while
buying an ETL product would prefer to use microsoft product instead of other different companies offering other ETL products. The other users also create various documentation which can be easily accessed. Even though the SSIS is a well developed product but it lacks the support for integration of different styles. One limitation of SSIS is its dependency on Microsoft environment. MS SQL server Integration Services can work with other product, but the integration between the applications is not trouble-free. Integration requires more time and effort which should be focused on other aspects.

VI. SUPPORT AND DOCUMENTATION/ARCHIVE

SSIS offers service level agreement and Mission control support. And on the other hand the Talend Do not offer SLAs and mission control. The SSIS provides much support and user guide documentation in case of situation where guidance is required. Talend does not offer Parallelization while the SSIS does offer parallelization. Talend also does not provide Dynamic schema support implementation where as the SSIS does provide the implementation. Data lineage in ETL process is provided by SSIS where as the Talend does not.

Data Quality

The data quality process includes cleansing of data, validating it, manipulation,doing quality tests, refining, and filtering of data. Most of the data quality tasks are not supported by Talend where as they are supported in SSIS.

VII. CONCLUSION

Both SSIS and Talend open Studios have their strengths and weakness. SSIS on the other hand performs better because of its stability, easy package configuration, parallel loading of data and good support. Though Talend is free but it is not usually suitable for commercial application where as SSIS is
more reliable. Considering the above mentioned advantages and disadvantages, Microsoft SSIS is far ahead when compared to Talend.

REFERENCES

[1] Retrieved from 'data warehousing' tutorialspoint.
[2] Retrieved from 'why Talend' https://www.talend.com/ home page .
[3] Retrieved from https://en.wikipedia.org/wiki/SQL_Server_Integration_Services.
[4] https://docs.microsoft.com/en-us/SQL/integration-services/SQL-server-integration-services.

.