Group – 14 | HW02 – Gartner Document | Parin Shah, Aishwarya Jonnalagadda, Shivam Verma, Shivam Mishra   
  
All the tools are prioritized according to their overall features and application towards the ETL process considering all the pros and cons is as follows: Key Features used to compare the tools are as below

1) Customer Service 2) Pricing 3) Product Capabilities & Service offerings 4) Integration & Deployment   
5) Bulk/Batch Data Movement 6) Platform Independence (on-premises, Inter-cloud, Multi-cloud)  
  
**1) Microsoft**- Microsoft seems the best solution compared to other solutions. It has a lower cost of ownership (TCO) and have versatile data integration functions in the SSIS import/export wizard. It is flexible to use SQL server/Azure data to integrate with different data sources. It has lots of services in Azure Stack, providing users a simple solution to deliver and analyze their data at scale. Microsoft tools are very good at Bulk/batch data movement through SSIS. SSIS provides the way to pull data in parallel using Sequence containers in control flow. You can design a package in such a way that it can pull data from non-dependent tables or files in parallel, which will help to reduce overall ETL execution time. Microsoft is very good at providing customer services and has received top rank for the part. The only thing Microsoft currently lacks is the Inter-cloud and Multi-cloud platform independence.   
  
**2) Oracle  
-** Oracle is a great tool for integration and cloud services. It is widely used for data replication, synchronization, and big data streaming. It offers multiple deployment options including cloud and hybrid to its customers for its integration tool portfolio. Since many companies use this platform, it is easy to get feedback and skilled individual for work. Big data SQL is good for data visualization. It has bulk/ batch integration capabilities. Speed of Oracle database is good even when the database is big or large. Oracle can manage multiple databases when the transaction is same by using its two-phase commit protocol. But it lacks in support and solution delivery. Also, it is comparatively expensive to other tools available in market for same job.  
  
**3) SAP**  
SAP BusinessObjects Data Integrator makes use of their advanced data profiling to ensure the quality and quantity of your enterprise data, which will help clients to understand the characteristics between tables in your source data. To filter out unwanted data based on your specific business rules it uses the data validation feature by building a firewall between the source and target systems. SAP shows the impact of data quality issues across all downstream systems or applications. SAP offers savings with lower deployment costs but needs more technical expertise than master data management teams. It provides extensively on-premise and on-cloud based applications. There is a difficulty for customers of SAP on not receiving professional answers to more complicated questions. SAP do not provide enough deep integration within product set.  
  
**4) Talend**   
- Talend is an open source data integration platform and ETL tool. It is flexible enough to work with all the data sources and platforms out there in the market. It primarily focuses on the transform part in the data integration. On the other hand, SSIS is particularly well suited for jobs that need to be consistent, repeatable, and error managed. Microsoft covers every part in the data integration process like reading data from different sources, analyzing, and cleansing data, running ETL processes and can handle migrations of large datasets. Whereas Talend Data Integration can perform bulk data migrations of up to a few million records. There is better data storage management and licensing options available in Microsoft services which makes it a better choice as compared to Talend.

*Gartner List of Data Integration Tools: Microsoft*

Parin Shah, 001306523

**What is ETL and Why it is important?**

ETLs are very important for Business Intelligence (BI). Data from different sources can be put into one data source and can be used to implement analytical solutions to derive insights.   
Main usage of ETLs : 1) Data Integration 2) Data Warehousing 3) Data Migration

Any size of the company can leverage the use of ETL to understand what’s going on in their processes and how they can leverage the data stored in the systems to drive business decisions.

**Microsoft**

- The ETL (*Extract-Transform-Load*) tool proposed by Microsoft is [SSIS (SQL Server Integration Services)](https://docs.microsoft.com/fr-fr/sql/integration-services/sql-server-integration-services?view=sql-server-2017), giving you an easy means of collecting, extracting and transforming your data and then making them actionable in a data warehouse system.

**Key Points of the Tool**

- SSIS as a Data Integration Service

- ADF as part of cloud data integration

**Strength**- Low ownership cost  
- It’s easy to use and implementation speed is faster  
- Integration with SQL Server/Azure data services  
- On premises data can be connected with Azure data  
- Service and Support  
- SSIS uses different data formats  
- It is easier to handle SSIS packages  
- Integration with VS Studio and Sequel server  
- Uses the SQL Server Destination instead of Object Linking and Embedding, Database; which can load data faster  
- Parallel data load is possible

**Weakness**- It’s less relevant to Inter-cloud or Multi-cloud, Works better with Microsoft ecosystem  
- Difficult to implement complex data integration   
- Low on Data Governance, Data Modelling, and Metadata Management  
- SSIS needs Management Studio rather than being published to reporting services to see package execution reports  
- If multiple packages are running simultaneously then it can create a problem.   
- SSIS uses high memory which can create a problem with SQL.   
- This could cause more CPU allocation which is because of parallel processing.

**References**  
<https://www.sarjen.com/ssis-advantages-disadvantages/>  
<https://www.gartner.com/reviews/market/data-integration-tools/compare>

*Gartner List of Data Integration Tools: Oracle*

*-*Shivam Mishra,001388030

Key Points & it’s purpose:

* Oracle offers several data integration products like Oracle golden gate Cloud services and Oracle Data Integrator Cloud Service.
* Golden gate of Oracle can be used for data replication, big data streaming and replication of data.
* Oracle data integrator is used for Bulk or batch integration.
* Big data SQL is good for data visualization.
* Oracle service bus is known for its messaging operations.
* For cloud services oracle has Integration Cloud service (iPaaS)

Strength:

* Oracle provides great data integration and data replication tools such as Golden Gate. It is also known for data synchronization capabilities.
* It provides microservice based architecture.
* Oracle has metered pricing option for many of its products, so you pay for what you use.
* Has strong portfolio and options for tools and services that can provision a HIP.
* Provides multiple deployment options including cloud and hybrid for integration tool portfolio.
* One of its biggest strength is that it has large partner networks which makes it easier for industries or companies to find skilled workers or service provider easily.
* Oracle provides lot of Industrial support strength for on-line backup and recovery and it has good software fault tolerance to prevent the disk failure.
* Oracle database speed is quite good even when the data base is large.
* Oracle has a superior ability to manage multiple databases within the same transaction using a two-phase commit protocol.
* Oracle is portable and can be used on various hardware platforms and also on many networking protocols.

Weakness:

* Oracle is expensive than most of its competitors providing similar facilities
* Support and service for client’s problem is one of the big issues.
* Its marketing strategy is not very good because of which the customers do not know about most of its modern integration capabilities.
* There is room for better integration and automated SQL tuning
* Initial setup of rules requires more technical expertise than master data management teams are typically used to.
* More connectivity options for Hadoop distributions
* Expensive for initial installations
* Business Objects many need several servers and extensive hardware.

References:

Magic\_quadrant\_for \_data\_integration pdf

<https://www.erpgreat.com/oracle-database/advantage-of-oracle-database>

https://www.trustradius.com/products/oracle-database/reviews?qs=pros-and-cons

***Gartner List of Data Integration Tools: SAP BODs (BusinessObjects Data Integrator)***

*-*Aishwarya Swetha Jonnalagadda,001054200

SAP is a Germany based company. It offers the data integration products such as SAP Data Services, SAP Replication Server, SAP Data Hub, SAP HANA, and SAP Cloud Platform Integration Suite. These services have almost eighty thousand customers (80,000).

Key Points & it’s purpose:

* BusinessObjects DI(BODs) is ETL tool and data integration.
* The main components of Data Integration are:
  + Job Servers.
  + Designer.
  + Along with these data unification, data profiling, data auditing and data cleansing.
* This tool helps us to take data from any source and Load it to any Warehouse.
* This tool supports the platforms such as Windows, Sun Solaris, AIX and Linux.
* This tool will also take care of loading the data and integrating them to different analytical environments and the same integrator in general used to build data warehouse, data marts etc.
* DI web administrator is an interface where it allows the clients to manage various repositories, web services, metadata and job servers.
* Scheduling jobs, monitoring them and batch execution can be taken care by this tool.

Strength:

* SAP Data Services provides exceptional functionalities for handling both structured and unstructured data and performs data integration, quality, and cleansing.
* Enterprise level solution for data processing of all structured/ unstructured data
* On-premises deployment
* Data quality and integration
* Supports ETL Migration
* Simplified maintenance
* Ability to access data of any size from any source through SAP data services
* Provides excellence in information management
* Helps to stay connected with third party application
* Deploy with faster time to market
* Deliver enterprise-class data integration performance.
* To get the efficient and best performance it uses the ability of SAP HANA.
* Gartner named SAP, a Leader in enterprise information management
* Transforming from data integration to business intelligence hand in hand is easy

Weakness:

* Integration with other products could be improved.
* Scheduling of jobs in the sequence is not available
* New connectivity options for cloud solutions must be added
* Debugging is not very easy. When one wants to see the flow of data from one Query to another, the process is not very smooth and efficient.
* Initial setup of rules requires more technical expertise than master data management teams are typically used to.
* More connectivity options for Hadoop distributions
* Expensive for initial installations
* Business Objects many need several servers and extensive hardware.

References:

www.businessintelligence.info/resources/assets/bo/dataintegrator.pdf

en.wikipedia.org/wiki/BusinessObjects\_Data\_Integrator

[www.softwaretestinghelp.com/best-etl-tools/](http://www.softwaretestinghelp.com/best-etl-tools/)

[www.guru99.com/sap-bods-tutorial.html](http://www.guru99.com/sap-bods-tutorial.html)

**TALEND** - Shivam Verma (001344145)

**Key Points & It’s purpose:**

* Talend is an open source data integration platform and ETL tool.
* Talend (headquartered in Redwood City) released its first commercial data integration software in year 2005 & Talend Open Studio for Data Integration (first product) in year 2006.
* Java is the primary language for developing Talend products and services.
* Talend provides services like data integration and data management under its Data Fabric product, it also supports data quality, cloud storage and Big Data under its different commercial products.
* Some products offered by Talend are: Talend Open Studio, Talend Data Fabric, Talend Data Management Platform, Talend Big Data Platform, Talend Data Services Platform, Talend Integration Cloud and Talend Stitch Data Loader.
* Talend’s data integration services helps organizations with managing and developing framework complexities by addressing both ETL for analytics and ETL for operational integration needs.
* Talend focuses on the transform and blending data part in the data integration phase.
* It lets users connect to a number of data bases. It is flexible enough to work with all the data sources and platforms out there in the market.
* Talend helps organizations with taking real-time decision making and promotes data driven decision making.
* Talend distributes the source code of its base modules under the Apache license.

**Strengths:**

* Unified platform is provided. Talend combines real time data, apps and process data integration with hybrid and multi-cloud support.
* Talend is supported by a colossal network of its users, it provides solutions to a lot of queries and questions.
* Talend also leverages a broad variety of open-source innovations, including Apache Beam, Spark and Kafka, which are used extensively in Talend Data Fabric.
* Talend has an excellent graphical user interface.
* Talend provides faster designing & pre-built widgets to integrate with databases and web services.
* Talend provide real-time integration with dashboards and incorporated control for quick sending over numerous hubs.
* Talend has a quality of technical support.
* It offers bidirectional metadata sharing and management.
* Useful for small data utilities and generates native code.
* Supports ETL and Data Migration.

**Weaknesses:**

* Limited Licensing options available, thus there is a demand for enterprise license models.
* Difficulties with migration between versions.
* Bugs are experienced. The Java platform in Talend is unstable and quirky sometimes.
* Lacks some administrative tools.
* Lack of proper training document.
* Low professional services satisfaction among customers

**References:**

[www.trustradius.com/products/talend-open-studio/reviews?qs=pros-and-cons](https://www.trustradius.com/products/talend-open-studio/reviews?qs=pros-and-cons)

[fr.wikipedia.org/wiki/Talend](https://fr.wikipedia.org/wiki/Talend)

[www.guru99.com/talend-tutorial.html](https://www.guru99.com/talend-tutorial.html)

[help.talend.com/reader/xVkdIonzSPLD~DZ85Z0A\_w/kMxNCQGZjb9Da6ZrNuiG2A](https://help.talend.com/reader/xVkdIonzSPLD~DZ85Z0A_w/kMxNCQGZjb9Da6ZrNuiG2A)