

# **Northeastern**

# **University**

**College of Engineering**



**INFO 7290: DATA WAREHOUSING AND BUSINESS INTELLIGENCE**

**Assignment: Week-2 Gartner Research**

**Team Members:**

- MAYANK MEHTA
- AVISH JADWANI
- ABHILASH HEMARAJ
- RAHIL KAPADIA

## **Gartner Research Tool: Talend**

**Description:** Talend is an Enterprise Data Integration tool founded in California, US. Which offers users across the public, private, and hybrid cloud. It is open-source software that lets you develop data pipelines across cloud technologies. Talend provides services like big data, data integration, data management, data excellence, data exploration, data preparation, and enterprise application integration.

**Features:** Talend Data Fabric, Talend Data Quality, Talend Data Catalog-API services, and Stitch Data Loader, Advanced data profiling

### **Pros:**

- This tool is easy to use, it has good compatibility with data sources, and the main advantage of the people is having it open source.
- This tool runs fast and long without any issues to work with millions of the documents at a one job run. It also provides predefined functionality to use complex java codes for developers.
- Works perfectly on big data technologies such as Sqoop, Hive, Spark, etc.
- Fraud pattern detection using Benford Law, Column set Analysis, Advanced matching analysis, and Time correlation Analysis, these industry-oriented features provided by the company.
- Also, the amazing feature is, we can use connectors such as Amazon S3, Elastic Search, Amazon SQS, EMAIL(SMTP), Bonita, JIRA, BOX MS Azure, Google cloud platform, and many more.

### **Cons:**

- Sometimes it slows the process of code generation and migration of millions of rows in a data warehousing environment.
- Spark streaming and to develop machine learning algorithms in Talend you need to subscribe real-time big data package
- There are some problems occurring on the Mac platform such as slowness of interfaces. For some tasks and processes, you may feel difficulty if you are not familiar with MSSQL products.
- It requires several hundreds of megabytes of RAM and storage to run applications smoothly for Windows users.
- Talend data fabric may sometimes cause problems, if users may face issues or unexpected results, it is very hard to debug and solve the problem.

## **Qlik (Attunity)**

**Description:** Qlik is a challenger in the magic quadrant. Qlik acquired Attunity in 2019 to enable companies to address the challenges of quickly transforming data into business value. Attunity's innovative data integration technology is the core of Qlik's platform that offers end-to-end data integration and analysis. Qlik offers a range of data integration products such as Qlik Replicate, Qlik Compose, Qlik Enterprise manager, and Qlik Data Catalyst with Qlik Enterprise being the most popular.

### **Pros:**

- Data Integration across all major platforms:

Qlik supports all major RDBMS and legacy systems, NoSQL, data warehouse, and Hadoop platforms on-premise and in the cloud.

- RDBMS: Oracle, SQL, Sybase, PostgreSQL
- Data Warehouse: Teradata, IBM Netezza, Vertica, Pivotal, Exadata
- Hadoop: Cloudera, MapR, Hortonworks
- Cloud: AWS, Azure, Google Cloud
- Messaging Systems: Apache Kafka
- Legacy Systems: RMS, VSAM, DB2 z/OS

Aside from platform support, Qlik has an intuitive GUI that automates data ingestion and replication. This makes Qlik the easiest to use and implement a data integration tool in the market.

- Automated Warehousing:

Unlike the traditional methods of building and managing data warehouses, Qlik's technology to automate ETL allows companies to reduce time, cost, and risk of business intelligence projects by reducing dependency on IT development resources.

- Scalability:

Qlik can efficiently process big data loads with parallel processing and assures that only the completed transactions within the time specified are processed. This allows to scale up as the data volume grows and support new uses cases for BI projects

### **Cons:**

- Pricing:

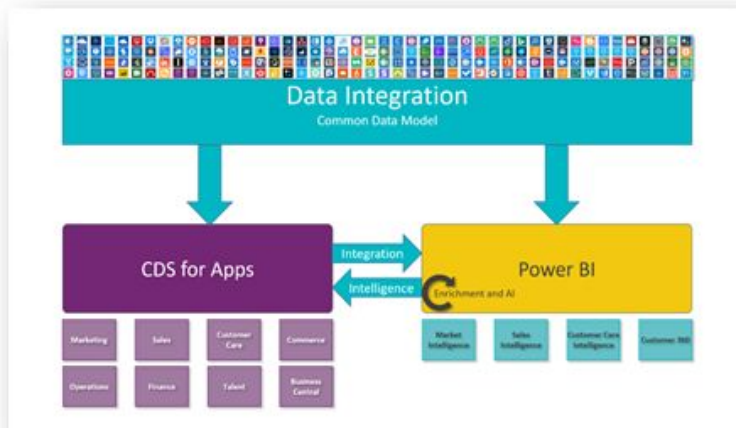
As Qlik offers a wide range of products for data integration and each product offers a different combination of servers, users, and applications-based licensing with a maintenance cost, customization cost, and change management fee, it is often difficult for companies to define their use case and select the right product

- Poor Customer Support:

According to reviews on the IT central portal, Qlik is often associated with poor technical assistance. Problems are usually related to deployment, user addition, and licensing of products.

## Data integration by Microsoft

**Description:** Microsoft provides data integration by ways of direct connectivity to data sources by bringing the data into Common Data Services (encapsulating apps from Marketing, Sales, Customer Care, etc. areas). It provides these services via SQL Server Integration Services (SSIS) and Azure data factory cloud integration services.



Some of the strongholds of SSIS:

- Provides connectivity via on-premises data gateway
- Integration into BI apps such as Power BI, Power apps, MS Flow, Azure Analysis Services, Power Query in Excel.

Above Image: [Data Integration Platform by Microsoft](#)

### Pros:

One of the strengths of Microsoft is its low TCO, and ease of usability. Coupled with SQL servers' on-prem data integration services with data stored in Azure services and its wide user base boasting one of the highest numbers of customers, makes Microsoft's DI services a fast-paced challenger amongst the other vendors. Microsoft is valued for its enhanced user support across all of its platforms participating in a hybrid system. Microsoft keeps top on its customer experience by ways of timely upgrades in its Integration platform. One of its recent upgrades includes AI support in Power Query, making way for BI's to gain access to AI transformations.

### Cons:

It has been observed by researchers at Gartner that Microsoft has a perception that their integration tooling must be used in a Microsoft centric development environment. Which is something which might cause them to lose relevance going forward. Even though Microsoft boasts for its ease of use, it is particularly complicated when it comes to data replication; This might cause customer drain as people might opt for alternatives for the same. Alongside this, Microsoft also suffers from poor data quality, data governance, and metadata management.

## **Oracle(Oracle Data Integrator):**

Oracle is a leader in the Data Integration market, providing several data integration tools both for cloud and on-premises platforms. Some of the tools are Oracle Data Integration Platform Cloud Service (includes Oracle GoldenGate Cloud Service and Oracle Data Integrator Cloud Service), Oracle Integration Cloud Service (iPaaS), Oracle Data Integrator (ODI — for bulk/batch integration), Oracle Big Data SQL (for data virtualization scenarios).

The Oracle Data Integration Cloud platform helps migrate and extract values from data by combining capabilities of Data Integration, Data Quality, and Data Governance into an individual unified cloud platform. The Oracle Data Integrator reduces the network traffic and uses the ELT architecture for data transformation.

The most common use cases include accelerate data warehouse, migrate data without any downtime, automate data mart generation, etc.

### **Pros:**

- Oracle Data Integrator supports multiple data sources for these tasks:
  - Databases like Oracle, SQL Server, Mysql
  - Kafka
  - Autonomous Data Warehouse
  - Flat files
- Oracle Integration Cloud service is the only competitor product to SAP's integration solution.
- It provides an easy-to-use user interface combined with a rich extensibility framework that helps Oracle Data Integrator improve productivity, reduce development costs, and lower the total cost of ownership for data-centric architectures.
- Rapid prototyping and development of integrations with the many out of the box adapters.
- Very promising on-premises integration with Software as a Service (SaaS), Platform as a Service(PaaS), Database Cloud Services(DBCS).
- A readily available developer pool, many organizations prefer to play around with a demo product before purchasing it.

### **Cons:**

- Performance issue for long-running threads.
- They need to work on the multiuser development environment and include the ability to comply with different kinds of SDLCs.

- ODI does not have a strong versioning process. This can create versioning issues between environments if you do not have a strong external code versioning process.

## **Recommendations:**

1. **Talend:** Has been a leader in the Data Integration sector. Performs highly on ease of usability. Addresses data quality and governance issues well.
2. **Microsoft:** Little to no scalability issues on the SSIS front. Suffers through the Azure Cloud factory platform since it being relatively new. Has a large user base.
3. **Oracle:** The licensing investment is expensive although, from a database security perspective, it is awesome. Holds an upper edge due to its contribution to cloud services and on-premises integration.
4. **Qlik:** Sometimes it is hard to modify the data after everything is set up that causes trouble for the ETL process but a lot of functionality

## **References:**

- <https://docs.microsoft.com/en-us/business-applications-release-notes/october18/data-integration-platform/overview>
- <https://www.gartner.com/en/documents/3955823/magic-quadrant-for-data-integration-tools>
- <https://www.itcentralstation.com/products/qlik-replicate-reviews>
- <https://www.capterra.com/p/171377/Attunity-CDC/>
- <https://docs.oracle.com/en/cloud/paas/data-integration-platform-cloud/using/what-is-dipc.html#GUID-CB0FE85B-7F74-4600-88CB-A598F73E891A>
- <https://www.ateam-oracle.com/integrating-oracle-data-integrator-odi-on-premise-with-cloud-services>