Write a report on the different ETL and ELT tools and technologies available in the market and the pros/cons for each.

ETL tools-

1. Hevo Data – Hevo Data is a Fully Automated, No-code Data Pipeline Platform that helps organizations leverage data effortlessly. Hevo’s End-to-End Data Pipeline platform enables you to easily pull data from all your sources to the warehouse and run transformations for analytics to generate real-time data-driven business insights.

Pros -

* + - Hevo is the best tool when you want to manage a large number of pipelines. With Hevo you can automate the pipelines very easily.
    - Hevo's great features are automatic detection and notification of failed data pipelines.
    - Auto mapping feature is also good. It's makes really easy to move the data.
    - Integration is easy with so many databases like SQL, Google Sheet, etc.

Cons -

* + - They should create a separate data source for SQL Server schema.
    - Some schema are not auto mapped, so need to do by manually.
    - Sometimes takes lot's of time to load the pipelines.Xplenty: Xplenty is a cloud-based ETL solution which requires no coding and provides simple visualized interface for performing ETL activities. It also connects with a large variety of data sources.

1. IBM – DataStage: It is a business intelligence tool for integrating data across various enterprise systems, it is part of IBM information platforms solution suite it uses visualized notation to make ETLprocesses, it is a powerful data integration tool.

Pros –

* Economical benefit – Elimination of CAPEX model, low operating cost
* Speed – It is faster to deploy a project without the need of setup
* Distributedness
* Reliability through redundancy
* Scalability
* Least burden of server sysadmin works
* Least one-time investment
* Least vendor lock-in
* Least downtime unlike normal server maintenance

Cons –

* Communication problems are possible resulting in poorer performance
* Ambivalence between security and ease of access to resources
* Higher cost for too bigger projects
* Lesser reach of product offerings, features to the client resulting in wrong product selection

1. Informatica: Informatica is leading market in data integration, Informatica’s suite of data integration software includes PowerCenter, which is known for its strong automation capabilities. Informatica PowerCenter is developed by Informatica Corporation. Informatica PowerCenter can connect to many sources for fetching data for data integration like PowerCenter Designer, Workflow Manager, Workflow Monitor, Repository Manager.

Pros –

* Informatica Powercenter is an innovative software that works with ETL-type data integration. Connectivity to almost all the database systems.
* Great documentation and customer support.
* It has a various solution to address data quality issues. data masking, data virtualization. It has various supporting tools or MDM, IDQ, Analyst, BigData which can be used to analyze data and correct it.

Cons -

* It doesn't have much scheduling options and the tool is not very much capable if we schedule more jobs.
* Debugging the workflows and mappings are very hard with Informatica PowerCenter.
* Lookup transformation on large tables consumes more Memory and CPU.
* Informatica PowerCenter have four client tools which is used in development process.

1. Microsoft SQL Server SSIS: Microsoft offers SSIS, a graphical interface for managing ETL using MS SQL Server. SSIS have user friendly interface, allowing users to deploy integrated data warehousing solutions without having to get involved with writing lots of code. SSIS is a fast and flexible data warehousing tool. The graphical interface allows for easy drag-and-drop ETL for multiple data types and warehouse destinations.

Pros –

* Provision of data integration services.
* Loading data warehouses with secure files for future use.
* Programming integration models without coding.

Cons -

* No room for improvement from my experience.
* Everything has been operating positively.

1. Talend:Talend is open source software that integrate, cleanse profile data and helps you get business insights easily. Talend has a GUI that enables managing a large number of source systems. This tool has Master Data Management(MDM) functionality. It also provides metadata repository using which user can easily re-use work.

Pros –

* Extract data from different type of sources - Great connectivity options available.
* Transform the data using multiple functions/components within Talend Data.
* Load the data into different types of target tables.
* One can also customize the code for specific use cases.

Cons-

* Scheduling options are limited.
* Using Machine Learning is not very easy.
* Documentation provided by Talend can be better.

1. Azure Data Factory: Microsoft Azure Data Factory is a cloud-based data integration service that automates the ETL process. We can say it is SSIS in the cloud because they share the same idea but SSIS provides more powerful GUI, debugging, and intelligence tools.

Pros –

* It allows copying data from various types of data sources like on-premise files, Azure Database, Excel, JSON, Azure Synapse, API, etc. to the desired destination.
* We can use linked service in multiple pipeline/data load.
* It also allows the running of SSIS & SSMS packages which makes it an easy-to-use ETL & ELT tool.

Cons -

* For complex JSON when it comes to mapping nested attribute it's not easy to flatten out
* Data Factory V1 does not have a good implementation experience as compared to V2
* Work with on premise solutions sometimes is not too friendly because you will need to set a VPN

1. Oracle Data Integrator: Oracle Data Integrator is based on Extract, load, and transform (ELT) architecture which means it performs load first and then transforms data. This tool is produced by Oracle which offers a graphical environment, and it is also very cost-effective.

Pros –

* + - The User Interface is pretty neat and clean which is very easy to use.
    - There are a lot of transformation options available to choose from.
    - The scalability and performance are top notch.

Cons -

* + - The cost part is bit high as compared to its peers.
    - Sometimes there is a lag, and it gets hanged.
    - Real time data Integration cannot be handled.

ELT tools-

1. Airbyte: Airbyte is an open-source ELT tool for modern data teams. It extracts raw data from sources such as databases and web applications with pre-built or custom connectors within minutes and loads them into the destination repositories with the Airbyte UI or APIs.

Each connector runs as a Docker container and allows you to use your language of choice. You and schedule pipeline updates.

Pros –

* Extensible: Using the Connector Development Kit, Data engineers can build custom connectors.
* Transformation: With DBT integration, engineers can design their custom transformations.
* Automation: You can automate your replication frequency and perform incremental updates to save costs.
* Support: Since it is an Open-Source ETL Tool, there is a growing community of developers that updates the connectors when the source APIs & schemas evolve.

Cons-

* To set up your data pipeline, you need technical knowledge and spend a significant portion of your engineering bandwidth to develop, maintain & manage your connectors.
* There are frequent new releases & lack of effective error handling.
* The UI and the tool are not beginner-friendly for non-technical users like business analysts.
* You need technical know-how to set up your custom transformations.
* You can only perform transformations after loading the data to your data warehouse. It lacks pre-load transformation capabilities.

1. StreamSets: StreamSets is a continuous big data loading platform that builds resilient data pipelines and manages data modification. It helps build smart data pipelines that cater to all data engineering life cycle needs. These pipelines facilitate data migration across hybrid and multi-cloud environments without requiring rewrites. StreamSets also enhances the ELT process by allowing data pipeline streaming for transparent data movement and management.

Pros-

* To set up your data pipeline, you need technical knowledge and spend a significant portion of your engineering bandwidth to develop, maintain & manage your connectors.
* There are frequent new releases & lack of effective error handling.
* The UI and the tool are not beginner-friendly for non-technical users like business analysts.
* You need technical know-how to set up your custom transformations.
* You can only perform transformations after loading the data to your data warehouse. It lacks pre-load transformation capabilities.

Cons-

* Monitoring/Visualization can be improvised and enhanced a lot (e.g. to monitor a Job to see what happened 7 days back with data transfer).
* The logging mechanism can be simplified (Logs can be filtered with "ERROR", "DEBUG", "ALL" etc but still takes some time to get familiar for understanding).
* Auto Scalability for heavy load transfer (Taking much time for >5 million record transfer from JDBC to ADLS destination in Avro file transfer).
* There should be a concept of creating Global variables which is missing.

1. Blendo: Improvado is a tool built by marketers, for marketers to get all their data into one place, in real-time, with automated dashboards and reports. The Improvado platform is specifically focused on the marketing dilemma and will connect to any marketing platform you need. The integrations run deep, pulling in granular data from the keyword and ad level, to allow you to see the complete picture.

Pros –

* Full support with a customer service rep included
* Marketing integrations are deep and granular, so you can see data at the keyword or ad level
* Ability to create custom metrics and map data across platforms
* Great for brands managing spend across multiple digital marketing channels

Cons –

* Some of the more granular features can be a bit complicated, but support is great about walking users through them.
* In order to get your dashboards and reports visualized in exactly the way you want, there may be some initial back and forth with your customer support rep.

1. Fivetran: Fivetran replicates applications, databases, events and files into a high-performance data warehouse, after a five minute setup. The vendor says their standardized cloud pipelines are fully managed and zero-maintenance.

Pros –

* Near real time Data replication
* Basic standardization
* Availability of a wide variety of ingestion plugins

Cons-

* Near real time Data replication
* Basic standardization
* Availability of a wide variety of ingestion plugins

1. Stitch: Stitch is an open-source, cloud-first ELT data integration platform owned by Talend. It’s a self-service ELT tool that automates data pipelines for rapid data flow.

It offers a simple user interface that enables single-click data ingestion, empowering analysts and business users to perform data transformations and analytics.

Pros –

* Automate data ingestion.
* Easy integration with many sources
* Extensible

Cons –

* Better error messages so you can determine what the problem is.
* Users feel the UI is not as friendly.