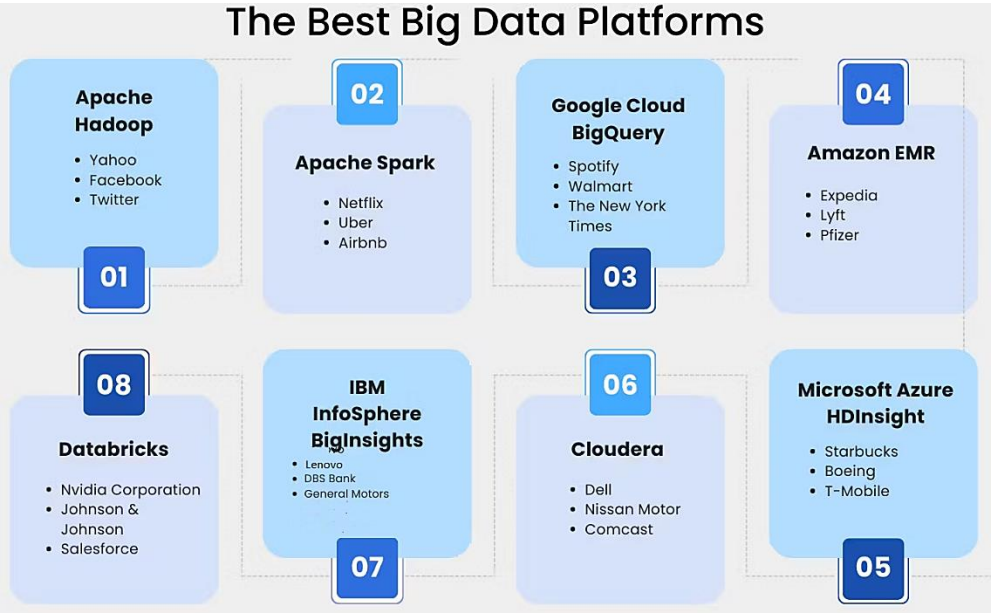
PRACTICAL-2

Perform the following tasks.

Task-1: Explore the various **trending Data analytics tools** available in the market and their functionalities and prepare a summary report.

Reference:



Example of some trending analytics tools: Hadoop, Spark, NoSQL databases (MongoDB, Cassandra), Hive, Pig, Flink, Kafka, HBase, Presto, Elasticsearch, Splunk, Tableau, Power BI, Talend, Apache NiFi, TensorFlow, RapidMiner, KNIME, DataRobot

Task-2: Explore any "TWO" real-world cases and understand the approach of problem solving in the Analytics World. (As per the Explanation) and summarize your understanding.

Reference: Netflix (OTT Platform)

Netflix, Inc. is a media company based in Los Gatos, California, founded in 1997 by American entrepreneurs Reed Hastings and Marc Randolph. The company has been pushing the envelope as a content deliverer since its inception. Netflix first disrupted the Blockbuster video-rental model; it then took on video streaming and cable companies, then content creators, and, as of 2023, broadcasters.

1997–2006: Video rental- Netflix was launched in 1997 as a DVD rental service that charged a per-rental fee. With tens of thousands of titles available, subscribers could select movies and TV shows through the website and receive the discs by mail. In 1999, Netflix offered an online subscription service; by the year 2000, this included flat-fee unlimited rentals with no due dates, late fees, or shipping fees. In 2000, the company introduced a personalized movie recommendation system, using an algorithm to predict an individual's movie preferences based on previous rental data. This system was improved after the Netflix Prize contest in 2006 challenged 40,000 contestants to improve the accuracy of movie recommendations by 10%. The $1 million Netflix Prize was awarded in 2009 to BellKor's Pragmatic Chaos, a team made up of seven mathematicians, computer scientists, and engineers from the United States, Canada, Austria, and Israel.

2007–2012: Streaming- In 2007, Netflix heralded a new era with the introduction of streaming services that allowed subscribers to access content directly over the Internet. Streaming delivery quickly gained momentum, becoming unlimited for most subscription plans. Netflix collaborated with makers of video game consoles, Blu-ray Disc players, and other electronics to enable video streaming on those devices. By 2010, Netflix introduced a streaming-only plan. The company initiated its global expansion to Canada, Latin America, the Caribbean, and Europe by 2012, reaching over 190 countries and territories by 2016.

2013–2020s: Original content- As streaming became its primary revenue generator; Netflix shifted its focus to the production of original content in 2013. By 2023, Netflix boasted more than 3,600 original titles. In 2022, Netflix announced plans to begin cracking down on password sharing, noting that over 100 million households had shared their account passwords with others. The crackdown was credited with boosting subscriber rates, with Netflix reporting it had added 8.8 million subscribers worldwide in the third quarter of 2023. The company also said it would raise prices on its streaming services. Netflix could be confident in cracking down on password sharing and raising prices because it sensed users would be willing to pony up for its broad collection of original programming.

2023–present: Live programming- In 2023, Netflix began live streaming with its stand-up special Chris Rock: Selective Outrage. In January 2024, Netflix said it had inked a 10-year deal with World Wrestling Entertainment (WWE) to stream a weekly live professional wrestling program, Monday Night Raw, beginning in 2025. In March 2024, the company announced it would live stream a fight between boxers Mike Tyson and Jake Paul in July.

Using big data at source, Netflix is decided to help the customers and provide the better valuable information using various analytic tools and methods. Following real world problems solved using big data:

Problem: How to suggest the movies/shows based on the user interest?
Solution: Recommendation based on interest
Problem: How to suggest the movies/shows based on their purchase?
Solution: Recommendation based on buying or browser history
Problem: How to increase the product sale?
Solution: Using various ML/AI methods

Task-3: A cloud-based platform called Databricks uses the open-source big data processing engine Apache Spark to manage and analyse massive datasets. It provides a centralised workspace where business analysts, developers, and data scientists may work together to create and implement data-driven applications. Databricks offers tools and services for real-time analysis, machine learning, and data preparation, with the goal of simplifying and streamlining the process of working with large amounts of data. Support for numerous data types, interaction with well-known data science frameworks and libraries, and flexibility in scaling up and down when necessary are some of Databricks' salient features.

Perform following:

- First sign-up and register on it using link: https://community.cloud.databricks.com/
- Login into: https://community.cloud.databricks.com/
- Understanding User-Interface.
- Learning How to upload Data, Cluster creation, and write the Basic script of Apache Spark.

References:

- https://www.decisivedge.com/casestudies/
- https://www.infosys.com/services/data-analytics/case-studies.html
- https://tataelxsi.com/design-digital/big-data-engineering
- https://www.bernardmarr.com/img/bigdata-case-studybook_final.pdf