Introduction to Tools and Environment

➢With the increasing demand for Data Analytics in the market, many tools have emerged with various functionalities for this purpose. Either open-source or user-friendly, the top tools in the data analytics market are as follows.

➢**R programming** – This tool is the leading analytics tool used for statistics and data modeling. R compiles and runs on various platforms such as UNIX, Windows, and Mac OS. It also provides tools to automatically install all packages as per user-requirement.

➢**Python** – Python is an open-source, object-oriented programming language which is easy to read, write and maintain. It provides various machine learning and visualization libraries such as Scikit-learn, TensorFlow, Matplotlib, Pandas, Keras etc. It also can be assembled on any platform like SQL server, a MongoDB database or JSON

Introduction to Tools and Environment

➢**Tableau Public** – This is a free software that connects to any data source such as Excel, corporate Data Warehouse etc. It then creates visualizations, maps, dashboards etc with real-time updates on the web.

➢**QlikView** – This tool offers in-memory data processing with the results delivered to the end-users quickly. It also offers data association and data visualization with data being compressed to almost 10% of its original size.

➢**SAS** – A programming language and environment for data manipulation and analytics, this tool is easily accessible and can analyze data from different sources.

➢**Microsoft Excel** – This tool is one of the most widely used tools for data analytics. Mostly used for clients' internal data, this tool analyzes the tasks that summarize the data with a preview of pivot tables.

Introduction to Tools and Environment

➢**RapidMiner** – A powerful, integrated platform that can integrate with any data source types such as Access, Excel, Microsoft SQL, Tera data, Oracle, Sybase etc. This tool is mostly used for predictive analytics, such as data mining, text analytics, machine learning.

➢**Apache Spark** – One of the largest large-scale data processing engine, this tool executes applications in Hadoop clusters 100 times faster in memory and 10 times faster on disk. This tool is also popular for data pipelines and machine learning model development.

Introduction to Tools and Environment

➤The future data analytics environment must expand to incorporate a full spectrum of analytics utilities and capabilities, including:

➤**Predictive analytics**, which uses data mining, machine learning and artificial intelligence techniques to develop models for predicting future behaviors.

➤**Prescriptive analytics**, which provides recommendations for optimal outcomes of selected options based on predictive analytics. In other words, prescriptive analytics helps automate decision processes.

➤**Integrated analytics**, which allows developed analytical models to be integrated within information flow to execute automated decision support and execution.

➤**Feature extraction and text analytics**, which helps automatically identify and extract features from semi-structured and unstructured data that can then be used to fuel predictive and prescriptive analysis.