

**Python:**

Python is a versatile programming language widely used in data science for its simplicity, readability, and extensive ecosystem of libraries and frameworks. Popular Python libraries like NumPy, Pandas, Matplotlib, and scikit-learn are used for data manipulation, analysis, visualization, and machine learning tasks.

**R:**

R is a programming language and environment specifically designed for statistical computing and graphics. It's commonly used in academia and research institutions for data analysis, statistical modeling, and visualization. R provides a wide range of packages for specialized statistical techniques and advanced data visualization.

**Apache Hadoop:**

Apache Hadoop is an open-source framework used for distributed storage and processing of large datasets across clusters of commodity hardware. It includes components like Hadoop Distributed File System (HDFS) for storage and MapReduce for parallel processing. Hadoop is commonly used for big data analytics, batch processing, and data warehousing applications.

**Apache Spark:**

Apache Spark is a fast and general-purpose cluster computing framework that provides in-memory data processing capabilities. It's designed for real-time and batch processing workloads and offers APIs in multiple languages like Scala, Java, and Python. Spark's core features include Spark SQL for querying structured data, Spark Streaming for real-time data processing, and MLlib for machine learning tasks.

**SQL (Structured Query Language):**

SQL is a standard programming language used for managing and querying relational databases. It's essential for data retrieval, manipulation, and analysis in various data science projects. Data scientists use SQL to extract insights from structured datasets stored in databases like MySQL, PostgreSQL, or SQL Server. Additionally, SQL is often integrated with other data processing frameworks for data transformation and aggregation tasks.