SACHU SAJI
ROLL NO : 46

Big data analytics tools

1. **Discovery Phase:**
   - **Tools:**
     - **Tableau:** Enables exploration and visualization of data for initial insights.
     - **QlikView:** Allows interactive data discovery and visualization.
     - **Google Analytics:** Provides insights into website and app usage.
2. **Data Preparation Phase:**
   - **Tools:**
     - **Trifacta:** Assists in cleaning and preparing raw data for analysis.
     - **Pandas (Python library):** Used for data manipulation and cleaning.
     - **OpenRefine:** Helps clean and transform messy data.
3. **Model Planning Phase:**
   - **Tools:**
     - **KNIME:** Supports the creation of data science workflows for model planning.
     - **Alteryx:** Provides a platform for data blending and advanced analytics.
4. **Model Building Phase:**
   - **Machine Learning Frameworks:**
     - **Scikit-learn (Python):** Offers a variety of machine learning algorithms.
     - **TensorFlow and PyTorch:** Deep learning frameworks for neural network-based models.
     - **RapidMiner:** An open-source platform for data science and machine learning.
   - **AutoML Tools:**
     - **H2O.ai:** Automates machine learning model selection and tuning.
     - **DataRobot:** Provides automated machine learning for model building.
   - **Big Data Tools:**

        - **MLlib (Spark):** Part of Apache Spark, it offers scalable machine learning algorithms.
        - **MLflow:** An open-source platform to manage the end-to-end machine learning lifecycle.

5. **Communication of Results Phase:**
    - **Tools:**
        - **Tableau:** Used for creating interactive and shareable dashboards.
        - **Microsoft Power BI:** Allows creating reports and dashboards for effective communication.
    - **Presentation Tools:**
        - **Microsoft PowerPoint:** Used for presenting findings and insights.
        - **Google Slides:** A cloud-based alternative for creating and sharing presentations.
6. **Operationalization Phase:**
    - **Tools:**
        - **Apache Airflow:** Automates and orchestrates complex data workflows, including model deployment.
        - **Docker and Kubernetes:** Used for containerization and orchestration of deployed models.
        - **AWS SageMaker or Azure ML:** Cloud-based platforms for deploying and managing machine learning models.

Tools Used Across Multiple Phases:

- **Python:** Used in various phases for data preparation, model planning, and building.
- **Jupyter Notebooks:** Provides an interactive environment for writing code, making it versatile across different phases.
- **SQL:** Used for data preparation and exploration, especially in the initial stages of discovery.

Explanation of a Tool: Tableau (Discovery and Communication of Results Phases):

- **Description:** Tableau is a powerful data visualization and business intelligence tool.
- **Discovery Phase:**

    - *Working:* Analysts connect to various data sources, explore data using drag-and-drop features, and create interactive dashboards for initial insights. Tableau allows users to quickly visualize patterns and trends in the data.

- **Communication of Results Phase:**
    - *Working:* After the analysis, Tableau is used to create visually appealing and interactive dashboards, reports, and presentations. The tool facilitates effective communication of the findings to both technical and non-technical stakeholders, aiding in decision-making