Title: Dog Breed Classification

Introduction:

The aim of this project is to develop an image classification system for automatically identifying and categorizing dog breeds. The system intends to assist pet owners in finding dogs that match their interests and lifestyle, thereby facilitating successful adoptions. Additionally, the project aims to enhance medical care and treatments by providing veterinarians with insights into the characteristics of different dog breeds.

What Would You Like to Implement:

The implementation involves building an image classification system using deep learning and machine learning techniques. The system will process a dataset consisting of 1,17,671 images of 10 different dog breeds. The objective is to achieve high classification accuracy and provide valuable insights into dog breed identification.

Proposed Architecture and Tools:

The proposed architecture includes the following components:

- Data Preprocessing: Use Python and OpenCV for resizing images and storing them in Hadoop Distributed File System (HDFS).

- Data Storage: Utilize MongoDB to store metadata, performance metrics, and other relevant information for efficient data management.

- Model Development: Employ deep learning and machine learning models in Python for classifying dog breeds.

- Distributed Computing: Utilize Apache Spark for distributed computing to handle large-scale data processing.

- Data Visualization: Utilize Tableau for visualizing and exploring classification results.
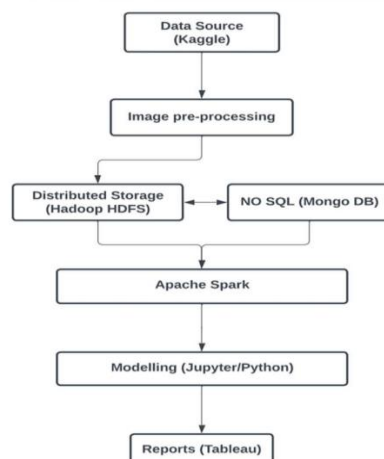
Plan of Work and Timeline:

1. Data Preprocessing (Week 1): Resize images and store them in HDFS. Set up MongoDB for metadata storage.

2. Model Development (Week 2-3): Develop and train deep learning and machine learning models for dog breed classification.

3. Distributed Computing (Week 4): Implement and optimize model training and evaluation on Spark cluster.

4. Data Visualization (Week 5): Create visualizations in Tableau to explore classification results and performance metrics.

5. Documentation and Reporting (Week 6): Prepare the final project report and presentation.

What Are You Planning to Achieve and Expected Results:

We aim to achieve:

- High classification accuracy for identifying dog breeds.

- Efficient handling of large-scale datasets (>100,000 images) using Hadoop and Spark.

- Visualization of classification results and performance metrics in Tableau.

Expected results include:

- Assessment of precision, recall, F1-score, and confusion matrices to evaluate classification quality.

- Good accuracy for large datasets (>100,000 images).

- Analysis of inference speed and latency for better system performance.

Git Account for the Implementation:

The implementation of the project will be available on the following Git repository:

GitHub Repository: https://github.com/Harsha2001-creater/603.git

Conclusion:

This intermediate report outlines the objectives, proposed architecture, timeline, initial findings, expected results, and Git repository for the implementation of the dog breed classification project. By leveraging deep learning, distributed computing, and data visualization techniques, we aim to develop an efficient and accurate system for identifying dog breeds from images.