1. Setting up hadoop on Single machine (standalone mode)

(Or)

Setting up hadoop on multi-node(distributed cluster environment)

1. Write a map-reduce program for word count in give text dataset.
2. Write a map-reduce program for finding maximum temperature in temperature dataset.
3. Write a map-reduce program for join of two records.
4. Write a map reduce program to find duplicate record in csv file.
5. Write a map reduce program to sum of numbers in text file.
6. Write a map reduce program to find patent citations in patent dataset.
7. Programs on retail dataset
8. Mapreduce program for total retail collection
9. Mapreduce program for store wise collection
10. Mapreduce program for product wise collection
11. Write a map reduce program for display highest ctc in each dept in employ ctc dataset.

Note: Cost to company (**CTC**) is a term for the total salary package of an **employee**