BIG DATA
DEVELOPMENT

ACADGILD

Session 4: MR -INTRODUCTION

Assignment 1

Big Data Hadoop and Spark Development

Assignment 1 – You must perform the given tasks.

Table of Contents

Big Data Hadoop and Spark Development

1. Introduction

In this assignment, you need to perform the given tasks.

2. Objective

This assignment will help you to consolidate the concepts learnt in the session 4.

3. Prerequisites:

None

4. Associated Data Files

https://drive.google.com/file/d/0Bxr27gVaXO5sVjQ5QW0wQ3RCTUU/view?usp=sharing

5. Problem Statement

We have a dataset of sales of different TV sets across different locations.

Records look like:
Samsung|Optima|14|Madhya Pradesh|132401|14200

The fields are arranged like:
Company Name|Product Name|Size in inches|State|Pin Code|Price

There are some invalid records which contain 'NA' in either Company Name or Product Name.

Task 1:

Write a Map Reduce program to filter out the invalid records. Map only job will fit for this context.

Task 2:

Write a Map Reduce program to calculate the total units sold for each Company.

Task 3:

Write a Map Reduce program to calculate the total units sold in each state for Onida company.

6. Expected Output

Solution report with commands, explanation to commands and screenshot for output.

Report shall be in PDF format. Submitted in GitHub.

7. Approximate Time to Complete Task

200 mins.