



BIG DATA
DEVELOPMENT

ACADGILD

Session 4: Exploring MapReduce 1

Assignment 2

Table of Contents

1. Introduction	3
2. Objective	3
3. Prerequisites:	3
4. Associated Data Files	3
5. Problem Statement	3
6. Expected Output	3
7. Approximate Time to Complete Task	3

Big Data and Hadoop Development

1. Introduction

In this assignment you need to perform the tasks given.

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.

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

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

6. Expected Output

None

7. Approximate Time to Complete Task