BIG DATA
DEVELOPMENT

ACAD**GILD**

Session 5: Getting Started with MapReduce

Assignment 1

Session 5: Assignment 1

Table of Contents

1. Introduction

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

2. 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

sales_differentTV :

https://github.com/prateekATacadgild/DatasetsForCognizant/blob/master/sales_differentTV.txt

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

Sales of different TV Task 1: Write a MapReduce program to filter out the invalid records. The output of this program will act as input for subsequent tasks. Map only job will fit for this context.

3. Output
N/A