

# Hadoop MapReduce Practice Workbook

Arjun Vankani

08-08-2025

## Introduction

This workbook contains a set of practice problems for Hadoop's MapReduce framework, ranging from simple counts to complex multi-stage data processing tasks. Each problem includes a complete solution outline to aid understanding.

## Problem 1: Log Level Count (Basic)

**Question:** You have a log file:

```
INFO: User A logged in
ERROR: Disk space low
INFO: User B uploaded file
ERROR: Connection timed out
INFO: User A logged out
```

Use MapReduce to count occurrences of each log level.

**Solution:**

**Mapping Phase:**

```
INFO, 1
ERROR, 1
INFO, 1
ERROR, 1
INFO, 1
```

**Shuffling:**

```
INFO -> (1, 1, 1)
ERROR -> (1, 1)
```

**Reducing:**

```
INFO, 3
ERROR, 2
```

## MapReduce Practice Problem 2

**Question:** Given the log file:

```
UserA: uploaded 2 files
UserB: uploaded 5 files
UserA: deleted 1 file
UserB: uploaded 3 files
UserC: uploaded 4 files
```

Use MapReduce to calculate the total number of files each user uploaded. Ignore delete operations.

## MapReduce Practice Problem 3

**Question:** From a web server log:

```
192.168.0.1 - GET /index.html
192.168.0.2 - GET /home.html
192.168.0.1 - GET /about.html
192.168.0.3 - GET /index.html
192.168.0.1 - GET /contact.html
```

Use MapReduce to find the number of requests per IP address and determine the most frequent visitor.

## MapReduce Practice Problem 4

**Question:** From a log file:

```
ERROR: Disk space low
ERROR: Disk read failure
WARNING: High memory usage
ERROR: Disk space low
WARNING: CPU temperature high
```

Use MapReduce to count each error/warning message type separately.

## MapReduce Complex Problem 5

**Question:** You are given a dataset of news articles with the following fields:

```
Date, ArticleText
2025-01-15, "Big data is the future of computing"
2025-01-20, "Hadoop processes big data efficiently"
2025-02-05, "Big data and AI are revolutionizing industries"
```

Write a MapReduce program to output the most frequent word for each month.