Election 2024 Analyzer Report

Task 02

Submitted by: Tazmeen Afroz
Roll Number: 22P-9252

Date: May 18, 2025

Task 1: Total Vote Casts

A MapReduce job was implemented to compute the total votes cast across all National and Provincial assembly seats.

Key-Value Pair Design

- **Mapper**:
    - **Key**: `Total Votes.`
    - **Value**: Vote count from each row .
- **Reducer**:
    - **Input**: `Total Votes.`
    - **Output**: Sum of all votes .

Questions

1. **Total Votes**: 114,032,235.
2. **Mapper Threads**: 7 mappers were launched.
3. **Reducer Threads**: 1 reducer was launched.
4. **Mapper Time**: 41,426 milliseconds (41.42 seconds).
5. **Reducer Time**: 2,760 milliseconds (2.76 seconds).
6. **Output File**: `hdfs:///usr/bda-22P-9252/task2/output1/part-0000.`

Task 2: Total Vote Casts (Variation 1)

This task analyzed execution times (mapper + reducer) by varying the block size of `election2024.csv` using `-D dfs.blocksize=<size>` and the number of reducer tasks using `-D mapred.reduce.task`

Execution Times

Table 1: Execution Times (seconds) by Reducer Tasks and Block Size

# Reducer Tasks	4 KB	8 KB	16 KB	32 KB	Default
2	30.838	31.856	30.953	31.578	30.814
4	35.430	36.058	35.880	34.040	36.255
8	48.552	47.280	47.006	45.687	45.729
16	70.737	73.401	73.671	74.508	72.914

Questions

1. **Number of Output Files for 16 Reducers**: 16 files, one per reducer.
2. **Zero-Byte Output Files**: None.

Task 3: Total Vote Casts (Variation 2)

This task analyzed execution times by varying the block size using `-D dfs.blocksize=<size>` and the number of mapper tasks using `-D mapred.map.tasks=<count>`.

Execution Times

Table 2: Execution Times (seconds) by Mapper Tasks and Block Size

# Map Tasks	4 KB	8 kb	16 KB	32 KB	Default
2	5.802	5.930	5.783	5.736	5.823
4	13.755	13.865	13.928	13.666	13.706
8	30.594	30.176	31.135	30.879	32.897
16	61.550	62.102	62.286	62.713	61.139

Task 4: Total Vote Casts (Variation 3)

using blocksize 4kb

Execution Times

Table 3: Execution Times (seconds) by Mapper and Reducer Tasks (Default Block Size)

# Map Tasks	2 Reducers	4 Reducers	8 Reducers	16 Reducers
2	7.812	11.762	22.211	48.703
4	15.927	19.391	30.898	58.952
8	32.978	37.556	48.115	76.394
16	75.460	79.380	96.928	126.725

Task 5: Designing Additional Queries

all the files output are also uploaded and are in the folder name **task5_outputs** (zip folder)

Task 5a: Total Votes Cast in Each Constituency

Output:

```
NA101    259009
NA103    260515
NA105    284957
...
NA265    124369
```

Key-Value Pair Design:

- **Mapper**:

    - **Key**: Constituency (e.g., `NA265`).

    - **Value**: Votes (integer, e.g., `305`).

- **Reducer**:

    - **Input**: `constituency`.

    - **Output**: `constituency total_votes` .(e.g., NA265 124369)

Task 5b: Total Votes Cast in National Assembly and Provinces

Output:

```
Khyber Pakhtunkhwa - 8136784
National - 58833128
Punjab - 35842278
Sindh - 11220045
```

Key-Value Pair Design:

- **Mapper**:

    - **Key**: Assembly type (e.g., `National`).

    - **Value**: Votes (integer, e.g., `305`).

    - **Example**: `National305`.

- **Reducer**:

    - **Input**: `assembly`.

    - **Output**: `assembly total_votes` (e.g., `National 58837129`).

Task 5c: Total Votes Cast Against Each Political Party

Output :

```
ANP      1277045
BAP      90817
IND      31995751
...
TLP      5759012
```

Key-Value Pair Design:

- **Mapper**:

- **Key**: Party (e.g., `IND`).

- **Value**: Votes (integer, e.g., `305`).

- **Example**: `IND3̂05`.

- **Reducer**:

- **Input**: `party`.

- **Output**: `party total_votes` (e.g., `IND 31995751`).

Task 5d: Total Number of Candidates in All Constituencies

Output Sample:

```
ANP      172
BAP      8
IND      9687
...
TLP      667
```

Key-Value Pair Design:

- **Mapper**:

- **Key**: Party (e.g., `IND`).

- **Value**: Candidate count (1 per row).

- **Example**: `IND1̂`.

- **Reducer**:

- **Input**: `party`.

- **Output**: `party total_candidates` (e.g., `IND 9687`).

Task 5e: Total Votes Cast Against Each Party in National/Provincial Assemblies

Output:

```
Khyber Pakhtunkhwa – ANP – 681259
Khyber Pakhtunkhwa – IND – 949299
...
Sindh – TLP – 328110
```

Key-Value Pair Design:

- **Mapper**:

- **Key**: Assembly type (e.g., `National`).

- **Value**: `party` (e.g., `IND3̂05`).

- **Example**: `National3̂05`.

- **Reducer**:
    - **Input**: `assembly`.
    - **Output**: `assembly – party – total_votes` (e.g., `National – IND – 24693046`).

Task 5f: Histogram of Total Votes Cast in All Constituencies

Output :

```
Votes between 0-25,000 – 3
Votes between 25,000-50,000 – 13
Votes between 50,000-75,000 – 48
Votes between 75,000-100,000 – 66
Votes between 100,000+ – 269
```

Key-Value Pair Design:

- **Mapper**:
    - **Key**: Constituency (e.g., `NA265`).
    - **Value**: Votes (integer, e.g., `305`).
    - **Example**: `NA265305`.
- **Reducer**:
    - **Input**: `constituency`.
    - **Processing**: Aggregates votes by constituency, bins into 25,000-vote ranges.
    - **Output**: `Votes between start-end – count` (e.g., `Votes between 0-25000 – 3`).

Task 5g: Histogram of Votes Cast for Top 3 Parties

Output :

```
IND – Votes between 0-10,000 – 9199
IND – Votes between 10,000-20,000 – 129
...
PML-N – Votes between 40,000+ – 295
```

Key-Value Pair Design:

- **Mapper**:
    - **Key**: Party (e.g., `IND`).
    - **Value**: `votes` (e.g., `3051`).
    - **Example**: `IND3051`.
- **Reducer**:
    - **Input**: `party`.

– **Processing**: Aggregates votes and counts, selects top 3 parties by candidate count, bins votes into 10,000-vote ranges.

– **Output**: `party - Votes between start-end - count` (e.g., `IND - Votes between 0-10000 - 9199`).