Find_Coolest_Hottest_Year_MapReduce

April 2, 2025

[1]:
```
import random
import csv

def generate_weather_data(filename="weather_data.csv", num_records=1000):
    years = list(range(2000, 2025))  # Years from 2000 to 2024
    locations = ["New York", "Los Angeles", "Chicago", "Houston", "Miami",␣
 ↪"Seattle"]

    with open(filename, mode="w", newline="") as file:
        writer = csv.writer(file)
        writer.writerow(["Year", "Temperature", "Location"])  # CSV Header

        for _ in range(num_records):
            year = random.choice(years)
            temperature = round(random.uniform(-10, 45), 1)  # Temperature in␣
 ↪Celsius
            location = random.choice(locations)
            writer.writerow([year, temperature, location])

    print(f"[INFO] Generated sample weather data in {filename}")

generate_weather_data()
```

```
[INFO] Generated sample weather data in weather_data.csv
```

[2]:
```
import csv
from collections import defaultdict

# Simulated Mapper Function
def mapper(data):
    """ Simulates the Mapper function by emitting (year, temperature) pairs. """
    mapped_data = []
    for line in data:
        year, temp, _ = line  # Extract Year, Temperature
        mapped_data.append((year, float(temp)))  # Convert temp to float
    return mapped_data

# Simulated Shuffle & Sort Phase
```

```python
def shuffle_and_sort(mapped_data):
    """ Simulates Hadoop's shuffle phase by grouping by key (year). """
    grouped_data = defaultdict(list)
    for year, temp in mapped_data:
        grouped_data[year].append(temp)
    return grouped_data

# Simulated Reducer Function
def reducer(grouped_data):
    """ Simulates the Reducer function by finding max and min temperature per␣
 ↪year. """
    reduced_results = {}
    for year, temps in grouped_data.items():
        reduced_results[year] = {"max": max(temps), "min": min(temps)}
    return reduced_results

# Main function to run the simulated MapReduce job
def run_mapreduce(file_path="weather_data.csv"):
    print("[INFO] Starting Simulated MapReduce Job...")

    # Read the data
    with open(file_path, "r") as file:
        reader = csv.reader(file)
        next(reader)  # Skip header
        data = list(reader)

    print(f"[INFO] Read {len(data)} records from {file_path}.")

    # Run Mapper
    mapped_data = mapper(data)
    print("[INFO] Mapper phase completed.")

    # Run Shuffle & Sort
    grouped_data = shuffle_and_sort(mapped_data)
    print("[INFO] Shuffle & Sort phase completed.")

    # Run Reducer
    reduced_results = reducer(grouped_data)
    print("[INFO] Reducer phase completed.")

    # Print Final Output
    print("\n[INFO] Final Output (Hottest and Coolest Year):\n")
    for year in sorted(reduced_results.keys()):
        print(f"Year: {year} | Hottest: {reduced_results[year]['max']}°C |␣
 ↪Coolest: {reduced_results[year]['min']}°C")

    print("\n[INFO] Simulated MapReduce Job Completed Successfully! ")
```

```
# Run the job
if __name__ == "__main__":
    run_mapreduce()
```

```
[INFO] Starting Simulated MapReduce Job…
[INFO] Read 1000 records from weather_data.csv.
[INFO] Mapper phase completed.
[INFO] Shuffle & Sort phase completed.
[INFO] Reducer phase completed.

[INFO] Final Output (Hottest and Coolest Year):

Year: 2000 | Hottest: 44.2°C | Coolest: -6.1°C
Year: 2001 | Hottest: 45.0°C | Coolest: -7.1°C
Year: 2002 | Hottest: 41.9°C | Coolest: -9.9°C
Year: 2003 | Hottest: 44.8°C | Coolest: -8.4°C
Year: 2004 | Hottest: 43.4°C | Coolest: -7.0°C
Year: 2005 | Hottest: 43.5°C | Coolest: -7.2°C
Year: 2006 | Hottest: 41.6°C | Coolest: -9.3°C
Year: 2007 | Hottest: 44.0°C | Coolest: -9.3°C
Year: 2008 | Hottest: 39.4°C | Coolest: -9.9°C
Year: 2009 | Hottest: 43.0°C | Coolest: -9.0°C
Year: 2010 | Hottest: 44.9°C | Coolest: -9.2°C
Year: 2011 | Hottest: 44.2°C | Coolest: -7.4°C
Year: 2012 | Hottest: 44.3°C | Coolest: -8.5°C
Year: 2013 | Hottest: 41.2°C | Coolest: -9.2°C
Year: 2014 | Hottest: 41.9°C | Coolest: -8.9°C
Year: 2015 | Hottest: 44.7°C | Coolest: -8.4°C
Year: 2016 | Hottest: 44.8°C | Coolest: -9.9°C
Year: 2017 | Hottest: 44.7°C | Coolest: -9.0°C
Year: 2018 | Hottest: 45.0°C | Coolest: -8.9°C
Year: 2019 | Hottest: 44.1°C | Coolest: -8.5°C
Year: 2020 | Hottest: 41.9°C | Coolest: -9.6°C
Year: 2021 | Hottest: 43.5°C | Coolest: -9.9°C
Year: 2022 | Hottest: 43.2°C | Coolest: -7.0°C
Year: 2023 | Hottest: 44.6°C | Coolest: -9.3°C
Year: 2024 | Hottest: 44.2°C | Coolest: -8.1°C

[INFO] Simulated MapReduce Job Completed Successfully!
```

[ ]: