Extreme Weathermen Challenge: Distributed Climate Analysis System

Problem Statement:

Build a scalable climate analysis engine to process CSV files from global weather stations.

Input:

- Directory: climate_data/
- Files: USW00094728_2000.csv etc.
- Columns: DATE,TMAX,TMIN,PRCP,SNOW,SNWD,QC_FLAG
- QC_FLAG indicates data quality. Invalid rows must be skipped.

Requirements:

1. Parallel CSV parsing and streaming processing.
2. For each station: hottest/coldest/wettest day, longest heatwave, yearly stats.
3. Global insights: top 10 extremes, decade trends.
4. Output: console, summary.json, climate_summary.sqlite

Edge Cases:

- Missing or conflicting rows
- Time consistency and leap years
- Handle millions of rows efficiently

Bonuses:

- CLI flags, regression analysis, colored output, Redis caching

Deliverables:

- Source code

- summary.json

- Run example: python main.py --start-year 2000 --region US --detect-trends