Capacity Estimation - Distributed Weather App

1. System Requirements

User Base Assumptions:

- Estimated **Daily Active Users (DAU)**: 10 million
- Peak traffic accounts for **2x DAU at peak hours**
- Users request weather updates **every 10 minutes**

API Requests Estimation:

- Average requests per user per day: **6** (every 10 minutes during active hours)
- Total daily API requests:
  = **60 million requests/day**
- Peak QPS (Queries per second):
  ≈ **695 QPS**
- Peak load multiplier: **2x** → Maximum peak QPS = **1400 QPS**

2. Storage Estimation

Weather Data Assumptions:

- Weather data per location stored as a **JSON object (~1KB)**
- Number of locations covered: **100,000** cities & towns
- Retention period for historical weather data: **30 days**
- Storage required per day: = **2.4GB/day**
- **Total storage for 30 days**:
  = **72GB**

3. Network Bandwidth Estimation

API Response Size & Data Transfer:

- Average API response size: **2KB**
- Daily data transfer:
  = **120GB/day**
- Peak bandwidth requirement:
  = **2.8MB/sec (~22.4 Mbps)**

4. Compute Resources Estimation

Backend Processing Power:

- **Average processing time per request:** 50ms
- **Required servers:**
  - Single-core CPU can handle **(1/50ms) = 20 requests/sec**
  - Servers needed: **(1400 QPS / 20) ≈ 70 servers** (considering redundancy)

5. Database Scaling Strategy

Read vs Write Distribution:

- Read-heavy workload (**95% reads, 5% writes**)
- **Database Choices:**
  - SQL (PostgreSQL) for structured data
  - NoSQL (Cassandra) for high-read scalability

Scaling Approach:

- **Sharding:** Distribute data across multiple DB instances
- **Replication:** Read replicas to handle high read traffic
- **Caching:** Redis for frequently accessed weather data

6. Load Balancing & Fault Tolerance

Load Balancing Strategy:

- **Global Load Balancer (CDN-based)** to distribute traffic across regions
- API Gateway to route requests efficiently

Fault Tolerance & Redundancy:

- **Multi-region replication** for disaster recovery
- **Auto-scaling** to handle traffic surges
- **Failover mechanisms** for high availability

7. Final Summary

- **Estimated Peak QPS:** 1400 QPS
- **Storage Needs:** ~72GB/month for historical data

- **Bandwidth Requirement:** 22.4 Mbps peak traffic
- **Compute Resources:** 70 servers for API processing
- **Database Scaling:** Replication + Sharding + Caching
- **Load Balancing:** CDN-based with Auto-Scaling