Use Case

An organization has deployed 100M weather sensors.

The goal is to gather the data transmitted from all devices in one database to do predictions, and trend analysis.

Main Data

Number of devices	100000000
Period (years)	10
Analysts	10

Assumptions

1. Data per hour is as good as per minute for trends

2. Still need to keep the data per minute for deeper analysis

Actor	CRUD Operations	Data in Operation	Op Type	Rate		Data Durability	Data Read/ Written	Data Life	Response Latency	Query Time	Data Read	Data Freshess
device	send metric data every minute	device_id, metrics	write	1666667	per s	one copy	1000 bytes	10 years				
ops	identify non-operational devices	device_id, metric times	read	1	per hr				1 hr	1 hr	Collection Scan	< 1 hour
ops	aggregate data per hour	device_id, metrics	write	1	per hr	majority		10 years				
analyst	run ~10 analytic queries per hour	temperature metrics	read	100	per hr				10 min	10 min	Collection Scan	< 1 hour

Operations

Actor	sensor device
Description	send weather data to the server
Operation Type	write
Data in Operation	device ID, time stamp, device metrics
Frequency	1.6 M/sec
	=100000000/60
Data size	1000 Bytes
Data life	10 years
Data durability	one node, no need to wait for majority

Actor	Data Scientist
Description	run analytic/trend query on temperature
Operation Type	read
Data in Operation	temperature metrics
Frequency	100/hour
	=10 scientists * 10 op/hr
Data read	collection scan
Data freshness	up to the last hour