NON-RELATIONAL DATABASE PRACTICUM: AIR QUALITY NOSQL DATABASE

KENT STATE
UNIVERSITY

MIS: 64082 DATABASE MANAGEMENT AND ANALYTICS
FALL - 2022

Name: Avinash Ravipudi
User Id: vravipud

TABLE OF CONTENTS

The Scenario:

The business scenario for this paper will be focused on monitoring air quality. Specially, how different types of pollutants that contribute to a decrease in air quality are collected and stored. Real-time data is used to monitor air quality, and it is updated and recorded.

One of the fundamental rights of human society is the right to breathe clean air. Since Hippocrates' treatise "On Airs, Waters, and Places" was penned around 400 B.C.E., the link between an individual's health and the quality of air has been recognized. Air pollution is currently a serious environmental concern that has a negative impact on human health. According to the World Health Organization (WHO), 9 out of 10 individuals inhaled polluted air in 2016, and more than half of the global population was exposed to levels of pollution that exceeded WHO's air quality criteria.

Monitoring pollutants is an essential parameter for sustainable living because of environmental change. Continuous monitoring will be beneficial to the population. There is a probability of surviving if the Quality level changes abruptly. Presently, the air quality monitoring system is considering the following pollutants for assessing the Air Quality: SO2, NO2, O3, and PM2.5, NO, BC (Black Carbon), PM10, CO. The contaminants are used to compute the air quality value. The data gathered and assessed to establish the degree of Quality is known as the Air Quality value. Every year, this information is gathered in the fall, spring, and summer. It is difficult to conceive how much data will need to be collected each year. The Air Quality departments must find a scalable storage solution for this massive volume of data.

Reason for NoSQL:

A NoSQL database is the ideal solution since the system must store a large amount of data collected by the Monitoring system. Because of their scalability, NoSQL databases are preferred over SQL databases. A SQL database will make it difficult for a system to maintain data if the types of contaminants used to assess air quality change in the future.

The following information is also considered: Temperature, Atmospheric Pressure, Humidity, Precipitation, Windspeed, Wind Direction, Solar Radiation, and UVI are some of the variables that may be monitored. When measuring air quality, several contaminants are considered, and some pollutants in the air have yet to be discovered. Only researchers can identify the most hazardous pollutants, but they are a technique of identifying new contaminants, which leads to a change in air quality. As a result, if this sort of information is gathered in the future, huge volumes of data must be saved consistently. JSON type documents and NoSQL data, rather than standard SQL, will be employed. It can store a variety of data types, including images and graphs. This type of data is tough to store in a SQL database.

NOSQL Form:

No SQL supports both document and key-value store models. This study will make use of AWS DynamoDB architecture and store the JSON data. This database system saves data from a continuous monitoring system. AWS IOT will be beneficial for this form of monitoring system, both financially and technically.

Point-in-time recovery is supported for both automated (continuous) backups and on-demand backups. It can store an incredibly vast quantity of data that can be easily searched. Though less adaptable in terms of where you may deploy it, DynamoDB integrates more easily with other AWS products and services. Key-value and range queries are available through query operation, while scanning offers another way to read data from a table.

DynamoDB is dependable and automatically backs up data in three distinct facilities.
For this work, only a few contaminants' information is gathered and shown in a database. The document will record the kind of pollutant as well as information about the pollutant. High-level details are solely used for data storage and exact measurement of Air Quality data. The following section will show the implementation of this system, as well as screen shots of the examples. Additional examples were constructed and can be included by request but were excluded from the paper to consolidate the information.

Implementation and Examples:

To store pollutants data, DynamoDB is employed, and a sample of six items were inserted to the database. In general, this data is continuously collected and monitored by an IOT system, which then shares it with the DynamoDB database for storage and calculation of Air Quality measurements. This data is collected and stored every minute because it changes. The object in DynamoDB will have information on the pollutant type, time of capture, and contaminant value. This is merely sample information that is acquired as a part of study and is saved in the database. Additionally, pictures of how DynamoDB stores pollutants data are provided.

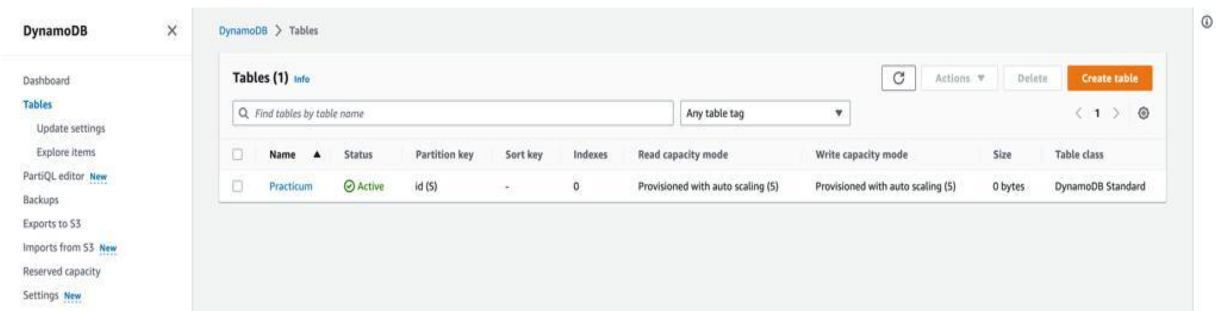

Image 1: Practicum Table creation in DynamoDB.

The database information and table name, "Practicum," are shown in the image above. The IOT system continuously adds pollutants to this database every single minute. The "id" in this case is regarded as the partition key. If we need to look for a specific pollutant, we can enter the Global Secondary Index as a pollutant. This paper solely covers write operations at this time. Global secondary Index can be introduced if read operations are taken into account.

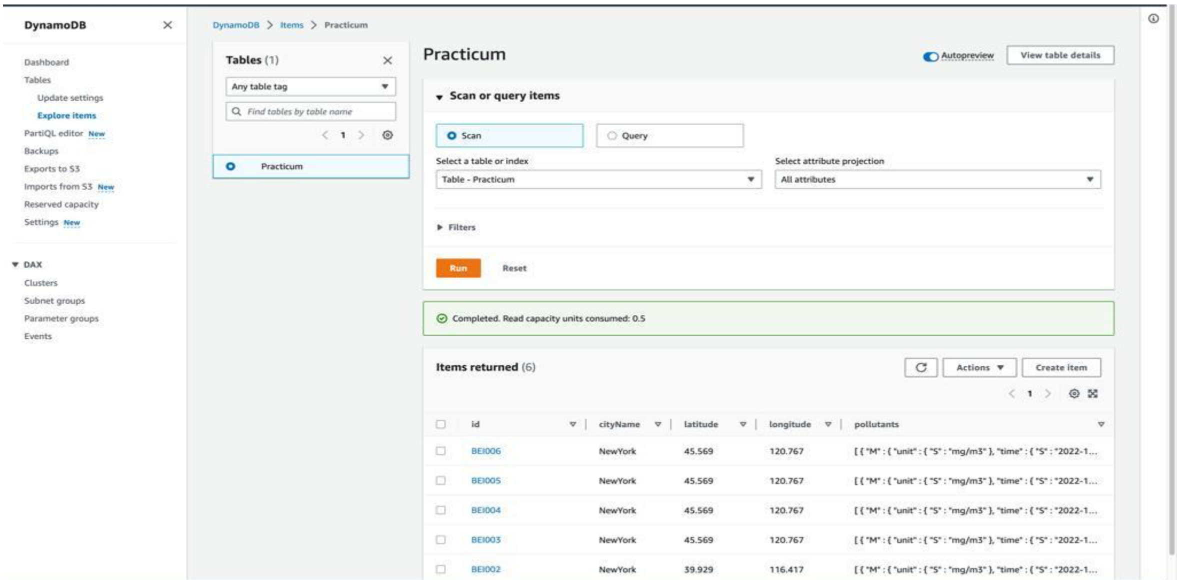

Image 2: Writing Items in Practicum DynamoDB Table.

Image 2 shows how the DynamoDB database looks like. Presently, there are six items that are being added and available in the system. As this data is real-time data, the items will get added into the database continuously. DynamoDB is scalable to store the huge amount of dynamic data.

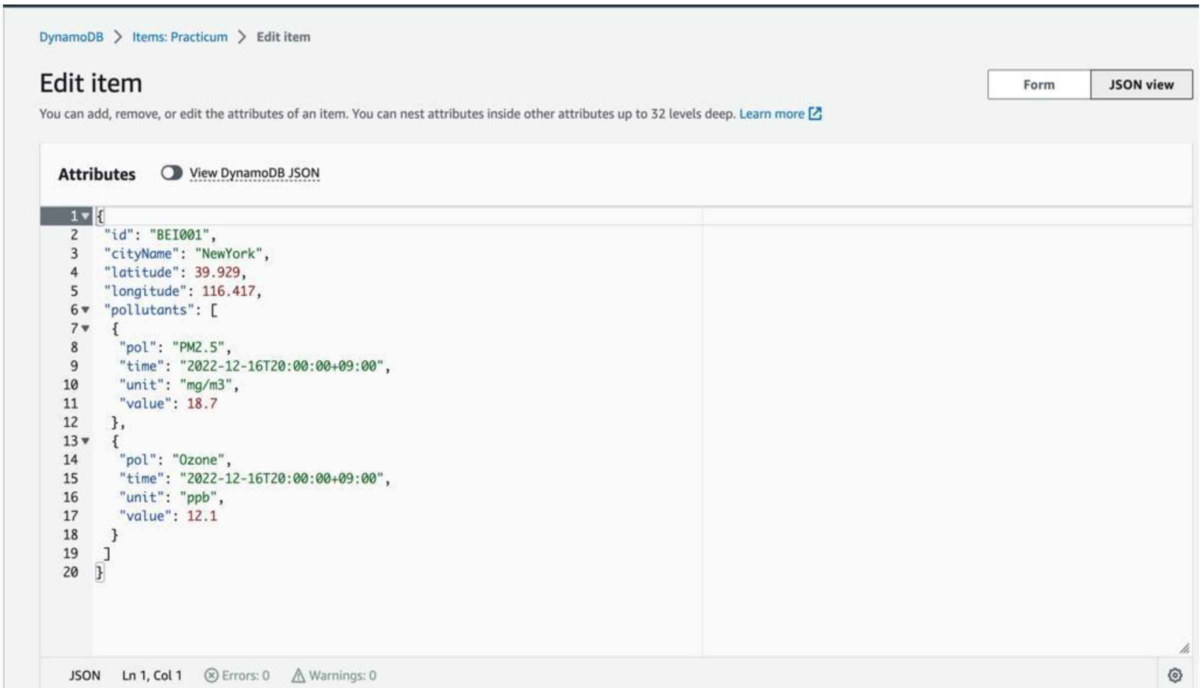

Image 3: Document for “BEI001” that includes basic information about the city New York and pollutants info at a certain time.

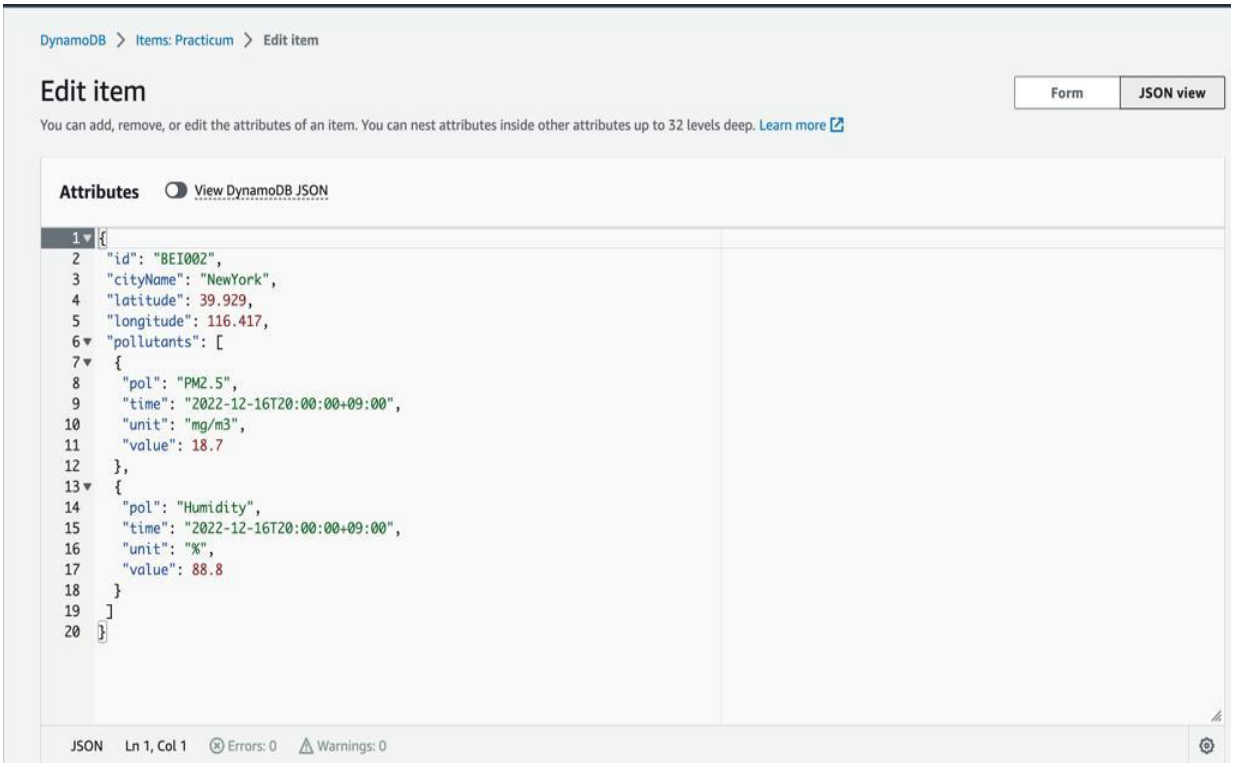

Image 4: Document for "BEI002" that includes basic information about the city New York and pollutants info at a certain time.

The images 3 and 4 above display various things that are noted as having different pollutants in the city of New York and values for the pollutants at a specific period that are accessible in DynamoDB.

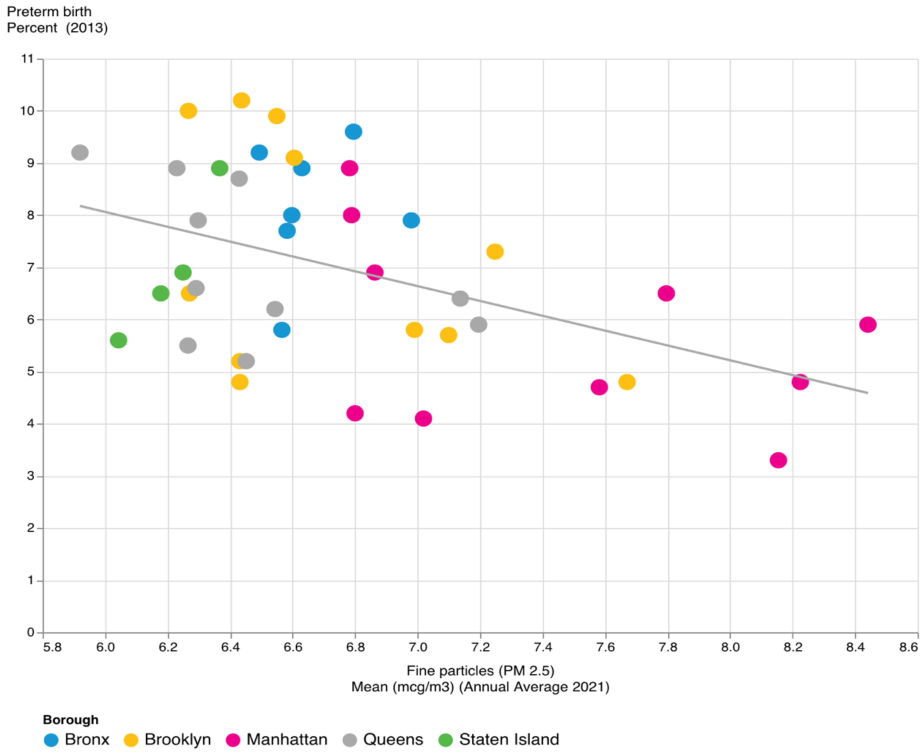

Image5: The graph displays the PM 2.5 pollutants data at different cities

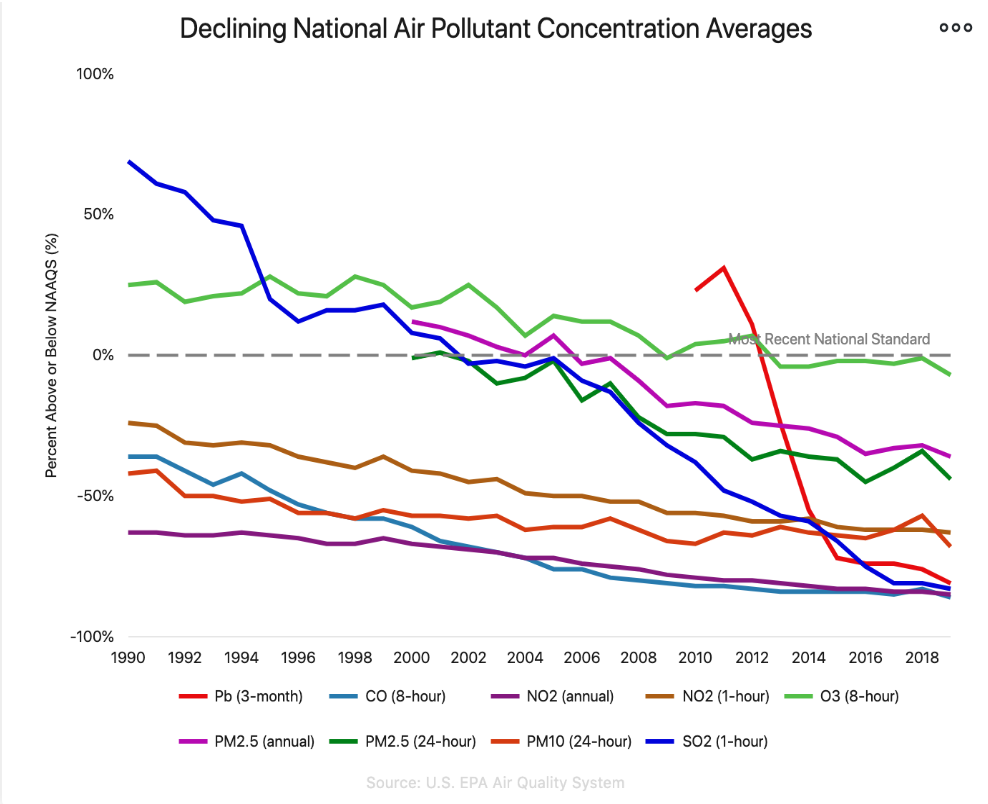

Image 6: Concentration Averages for different Air Pollutants

The data imported from the IOT system can be used to calculate a wide range of metrics. When tracking the Air Quality, some parameters, such those in Images 5 and 6, should be taken into account. The information above is presented as a whole rather than focusing on and identifying each individual pollutant in detail.

References:

1. Guide to publishing Air Quality data on the World Air Quality Index project
2. Research of time series air quality data based on exploratory data analysis and representation
3. Study Case: The Database Selection Process for the Big Data-based System to Reduce Health Effects of Air Pollution in Ciudad Juárez, Mexico.
4. The application of NoSQL database in Air Quality Monitoring
5. https://www.clarity.io/blog/air-quality-around-the-world-how-2020-served-as-a-global-laboratory-for-air-pollution-research