***Assignment 1***

Group Members-

|  |  |
| --- | --- |
| 19BIT087 | Nidhay Pancholi |
| 19BIT115 | Shambhavi Kumar |
| 19BIT088 | Nidhi Zala |
| 19BIT085 | Nandini Bhuva |
| 19BIT131 | Vivek Devre |

# 

**Contents**

[**The DBMS vs File system**](#_heading=h.1z2f6cvm4nxt) **2**

[DATA REDUNDANCY](#_heading=h.sc9cr7v1ms9r) 2

[DATA INCONSISTENCY](#_heading=h.3znysh7) 2

[EASY ACCESS TO DATA](#_heading=h.vqpg1rhwgm4n) 3

[DATA ISOLATION](#_heading=h.up1dkzpwmddy) 3

[DATA INTEGRITY](#_heading=h.3qy3v7swj7ge) 3

[ATOMICITY OF UPDATES](#_heading=h.q7zso68pmlr6) 4

[SECURITY](#_heading=h.4d34og8) 4

[CONCURRENT ACCESS](#_heading=h.nh8erh8tn8fl) 4

# 

# 

# 

# 

# The DBMS vs File system

## 

## DATA REDUNDANCY

1. Data redundancy occurs when same piece of data exists at several points or places at a time. Same data is stored at multiple locations.
2. The air quality data is collected at the various stations. They store their own data and also share it with the centres at the state levels. This data is again shared by the state at the national level.
3. If the need arises that some other department has to access the same data, the same data has to be shared all over again.
4. Sensors at the stations keeps on updating the data almost every minute, which leads to generation of huge amount of data.
5. If any private institutions or individual required data for research purpose then, the data has to be sent to them individually.

## DATA INCONSISTENCY

1. Data at different places is supposed to be the same but due to negligence or error the data differs from each other.
2. If the data of the sensors are being stored in the centralised system, then this is easier to manage.
3. For example, today’s data set has been sent to the state level but while forwarding it to the national station, the state level in charge sends yesterdays’ data. Now because of this the data at the national level for the respective state would be duplicated for two consecutive days.
4. For each sensor a different table has to be created when implementing a file system.
5. If some failure occurs while sending the files from station to the state level then the files have to be shared all over again.

## 

## 

## EASY ACCESS TO DATA

1. The sensors generate a lot of data as it is working every minute.
2. The data will be updated every minute which would be difficult to be analysed manually.
3. Hence if any GUI System is created for visualising and analysing data which can work with centralised database and the memory is easily accessible.
4. The data can be shared through API directly with the institutions or individuals for research purpose.
5. Various insights will be available to us because of easy access which would help in stopping the deteriorating conditions of the environment.

## 

## DATA ISOLATION

1. The latest updated data will not be available.
2. In case a programmer writes a code, then he’ll not get receive the latest data rather he’ll have to work with stale data.
3. If there is a centralised data, then a single code can be created to analyse data from all the different levels and stations.
4. Programmers at different levels will have to write their own individual codes.
5. The files at the station level may have different format, then there will be a problem in processing the data.

## 

## DATA INTEGRITY

1. The government can be a main cause of data being hampered. They can alter the data to demonstrate a clean image in front of public.
2. If DBMS is used then, this can surely be avoided and the data integrity can be maintained.
3. As we know that the sensors are the one going to write the data and hence all users should only be granted a read access.
4. In file system any one can modify the data for whatever reasons.
5. The concept of foreign keys cannot be implemented in file system and hence the data from various stations if necessary cannot be grouped together based on date, time, etc. Example : If every hour we want to calculate city and state averages for the concentration levels the data from all stations has to be grouped and then it has to be calculated.

## 

## 

## 

## 

## ATOMICITY OF UPDATES

1. If the sensors have got into some malfunctioning, then Null data will be updated in the file system continuously which cannot be altered.
2. This would create a serious problem in analysis system. If suppose the ozone sensors has stopped working, then the file system will receive Null values till the time it is repaired. This would alter the data analysis.
3. Hence if database is used then a condition can be put up such that if a sensor returns NULL values then those should be discarded and some other values should be substituted instead using some statistical analysis, based on time and location a time series analysis can be performed.
4. If any transaction has just been completed and power failure occurs, still the transaction has been saved. But if a power failure occurs while writing the data in a file then it has to be saved first to save the transactions permanently.

## SECURITY

1. In case of file system, the data will be stored in separate computers in various stations. That computer will be having several users. Hence, data can be easily hampered or altered.
2. Also, the computers can be hacked easily. Hundreds of computers cannot be provided with high security.
3. In DBMS the central system will be providing high security. All the data will be controlled using single security system and the legitimacy of data can be thus maintained.
4. No user needs to have an update access on the data.
5. In a database system all stations can view only the data shared by their respective station and same with the state level. For accessing the whole database they too would have to use an API.

## 

## CONCURRENT ACCESS

1. Several sensors can write data at the same time in DBMS system. This is not possible in file system. In that case we’ll have to compile system files separately.
2. In DBMS system all the sensors can write the data at same place and at the same time. The data from all the stations, states and nations can be combined, stored and accessed from this system itself. Whereas in case of file system the data has to be individually combined at different levels and then be shared.
3. The data stored in the file system will be available to only a few people. If some other person wants access to it, then he’ll have to undergo long procedures in order to receive the data.
4. If we create an API of the dataset, then the users throughout the nation can easily access it and get latest updated data.