QL | DDL, DQL, DML, DCL and TCL Commands

Structured Query Language(SQL) as we all know is the database language by the use of which we can perform certain operations on the existing database and also we can use this language to create a database. SQL uses certain commands like Create, Drop, Insert etc. to carry out the required tasks.

These SQL commands are mainly categorized into four categories as:

1. DDL – Data Definition Language
2. DQl – Data Query Language
3. DML – Data Manipulation Language
4. DCL – Data Control Language

Though many resources claim there to be another category of SQL clauses **TCL – Transaction Control Language**. So we will see in detail about TCL as well.

1. **DDL(Data Definition Language) :** DDL or Data Definition Language actually consists of the SQL commands that can be used to define the database schema. It simply deals with descriptions of the database schema and is used to create and modify the structure of database objects in the database.
   Examples of DDL commands:
   - **CREATE** – is used to create the database or its objects (like table, index, function, views, store procedure and triggers).
   - **DROP** – is used to delete objects from the database.
   - **ALTER**-is used to alter the structure of the database.
   - **TRUNCATE**–is used to remove all records from a table, including all spaces allocated for the records are removed.
   - **COMMENT** –is used to add comments to the data dictionary.
   - **RENAME** –is used to rename an object existing in the database.
2. **DQL (Data Query Language) :**
   DML statements are used for performing queries on the data within schema objects. The purpose of DQL Command is to get some schema relation based on the query passed to it.

   Example of DQL:
   - **SELECT** – is used to retrieve data from the a database.

3. **DML(Data Manipulation Language) :** The SQL commands that deals with the manipulation of data present in the database belong to DML or Data Manipulation Language and this includes most of the SQL statements.
   Examples of DML:
   - **INSERT** – is used to insert data into a table.
   - **UPDATE** – is used to update existing data within a table.
   - **DELETE** – is used to delete records from a database table.
4. **DCL(Data Control Language) :** DCL includes commands such as GRANT and REVOKE which mainly deals with the rights, permissions and other controls of the database system.
   Examples of DCL commands:
   - **GRANT**-gives user's access privileges to database.
   - **REVOKE**-withdraw user's access privileges given by using the GRANT command.
5. **TCL(transaction Control Language) :** TCL commands deals with the transaction within the database.
   Examples of TCL commands:
   - **COMMIT**– commits a Transaction.
   - **ROLLBACK**– rollbacks a transaction in case of any error occurs.
   - **SAVEPOINT**–sets a savepoint within a transaction.
   - **SET TRANSACTION**–specify characteristics for the transaction.

This article is contributed by **Dimpy Varshni**. If you like GeeksforGeeks and would like to contribute, you can also write an article using contribute.geeksforgeeks.org or mail your article to contribute@geeksforgeeks.org. See your article appearing on the GeeksforGeeks main page and help other Geeks.
Please write comments if you find anything incorrect, or you want to share more information about the topic discussed above.

Example of Database : Bigtable

A Distributed Storage System for Structured Data
Bigtable is a distributed storage system (built by Google) for managing structured data that is designed to scale to a very large size: petabytes of data across thousands of commodity servers.

Many projects at Google store data in Bigtable, including web indexing, Google Earth, and Google Finance. These applications place very different demands on Bigtable, both in terms of data size (from URLs to web pages to satellite imagery) and latency requirements (from backend bulk processing to real-time data serving).

Despite these varied demands, Bigtable has successfully provided a flexible, high-performance solution for all of these Google products.

Some features

- fast and extremely large-scale DBMS
- a sparse, distributed multi-dimensional sorted map, sharing characteristics of both row-oriented and column-oriented databases.
- designed to scale into the petabyte range
- it works across hundreds or thousands of machines
- it is easy to add more machines to the system and automatically start taking advantage of those resources without any reconfiguration
- each table has multiple dimensions (one of which is a field for time, allowing versioning)
- tables are optimized for GFS (Google File System) by being split into multiple tablets - segments of the table as split along a row chosen such that the tablet will be ~200 megabytes in size.

Architecture

BigTable is not a relational database. It does not support joins nor does it support rich SQL-like queries. Each table is a multidimensional sparse map. Tables consist of rows and columns, and each cell has a time stamp. There can be multiple versions of a cell with different time stamps. The time stamp allows for operations such as "select 'n' versions of this Web page" or "delete cells that are older than a specific date/time."

In order to manage the huge tables, Bigtable splits tables at row boundaries and saves them as tablets. A tablet is around 200 MB, and each machine saves about 100 tablets. This setup allows tablets from a single table to be spread among many servers. It also allows for fine-grained load balancing. If one table is receiving many queries, it can shed other tablets or move the busy table to another machine that is not so busy. Also, if a machine goes down, a tablet may

be spread across many other servers so that the performance impact on any given machine is minimal.

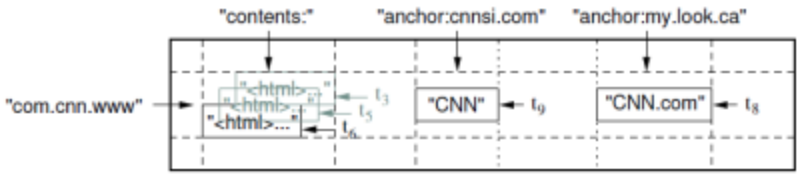
Tables are stored as immutable SSTables and a tail of logs (one log per machine). When a machine runs out of system memory, it compresses some tablets using Google proprietary compression techniques (BMDiff and Zippy). Minor compactions involve only a few tablets, while major compactions involve the whole table system and recover hard-disk space.

The locations of Bigtable tablets are stored in cells. The lookup of any particular tablet is handled by a three-tiered system. The clients get a point to a META0 table, of which there is only one. The META0 table keeps track of many META1 tablets that contain the locations of the tablets being looked up. Both META0 and META1 make heavy use of pre-fetching and caching to minimize bottlenecks in the system.

Implementation
BigTable is built on **Google File System** (GFS), which is used as a backing store for log and data files. GFS provides reliable storage for SSTables, a Google-proprietary file format used to persist table data.
Another service that BigTable makes heavy use of is **Chubby**, a highly-available, reliable distributed lock service. Chubby allows clients to take a lock, possibly associating it with some metadata, which it can renew by sending keep alive messages back to Chubby. The locks are stored in a filesystem-like hierarchical naming structure.
There are three primary **server types** of interest in the Bigtable system:

1. Master servers: assign tablets to tablet servers, keeps track of where tablets are located and redistributes tasks as needed.
2. Tablet servers: handle read/write requests for tablets and split tablets when they exceed size limits (usually 100MB - 200MB). If a tablet server fails, then a 100 tablet servers each pickup 1 new tablet and the system recovers.
3. Lock servers: instances of the Chubby distributed lock service. Lots of actions within BigTable require acquisition of locks including opening tablets for writing, ensuring that there is no more than one active Master at a time, and access control checking.

Example from Google's research paper:



A slice of an example table that stores Web pages. The row name is a **reversed URL**. The contents column family contains the **page contents**, and the anchor column family contains the **text of any anchors** that reference the page. CNN's home page is referenced by both the Sports Illustrated and the MY-look home pages, so the row contains columns
named anchor:cnnsi.com and anchor:my.look.ca. Each anchor cell has **one version**; the contents column has **three versions**, at timestamps t3, t5, and t6.

API

Typical operations to BigTable are creation and deletion of tables and column families, writing data and deleting columns from a row. BigTable provides this functions to application developers in an API. Transactions are supported at the row level, but not across several row keys.

What is Open Data?

In simple terms, Open Data means the kind of data which is open for anyone and everyone for access, modification, reuse, and sharing.
Open Data derives its base from various “open movements” such as open source, open hardware, open government, open science etc.

Governments, independent organizations, and agencies have come forward to open the floodgates of data to create more and more open data for free and easy access.

Why Is Open Data Important?

Open data is important because the world has grown increasingly data-driven. But if there are restrictions on the access and use of data, the idea of data-driven business and governance will not be materialized.
Therefore, open data has its own unique place. It can allow a fuller understanding of the global problems and universal issues. It can give a big boost to businesses. It can be a great impetus for machine learning. It can help fight global problems such as disease or crime or famine. Open data can empower citizens and hence

can strengthen democracy. It can streamline the processes and systems that the society and governments have built. It can help transform the way we understand and engage with the world.

Open Data sources:

1. World Bank Open Data

As a repository of the world's most comprehensive data regarding what's happening in different countries across the world, World Bank Open Data is a vital source of Open Data. It also provides access to other datasets as well which are mentioned in the data catalog.

World Bank Open Data is massive because it has got 3000 datasets and 14000 indicators encompassing microdata, time series statistics, and geospatial data.

Accessing and discovering the data you want is also quite easy. All you need to do is to specify the indicator names, countries or topics and it will open up the treasure-house of Open Data for you. It also allows you to download data in different formats such as CSV, Excel, and XML.

If you are a journalist or academic, you will be enthralled by the array of tools available to you. You can get access to analysis and visualization tools that can bolster your research. It can felicitate a deeper and better understanding of global problems.

You can get access to the API which can help you create the data visualizations you need, live combinations with other data sources and many more such features.

Therefore, it's no surprise that World Bank Open Data tops any list of Open Data sources!

2. WHO (World Health Organization) — Open data repository

WHO's Open Data repository is how WHO keeps track of health-specific statistics of its 194 Member States.

The repository keeps the data systematically organized. It can be accessed as per different needs. For instance, whether it is mortality or burden of diseases, one

can access data classified under 100 or more categories such as the Millennium Development Goals (child nutrition, child health, maternal and reproductive health, immunization, HIV/AIDS, tuberculosis, malaria, neglected diseases, water and sanitation), non communicable diseases and risk factors, epidemic-prone diseases, health systems, environmental health, violence and injuries, equity etc.

For your specific needs, you can go through the datasets according to themes, category, indicator, and country.

The good thing is that it is possible to download whatever data you need in Excel Format. You can also monitor and analyze data by making use of its data portal.

The API to the World Health Organization's data and statistics content is also available.

3. Google Public Data Explorer

Launched in 2010, Google Public Data Explorer can help you explore vast amounts of public-interest datasets. You can visualize and communicate the data for your respective uses.

It makes the data from different agencies and sources available. For instance, you can access data from World Bank, U. S. Bureau of Labor Statistics and U.S. Bureau, OECD, IMF, and others.

Different stakeholders access this data for a variety of purposes. Whether you are a student or a journalist, whether you are a policy maker or an academic, you can leverage this tool in order to create visualizations of public data.

You can deploy various ways of representing the data such as line graphs, bar graphs, maps and bubble charts with the help of Data Explorer.

The best part is that you would find these visualizations quite dynamic. It means that you will see them change over time. You can change topics, focus on different entries and modify the scale.

It is easily shareable too. As soon as you get the chart ready, you can embed it on your website or blog or simply share a link with your friends.

4. Registry of Open Data on AWS (RODA)

This is a repository containing public datasets. It is data which is available from AWS resources.

As far as RODA is concerned, you can discover and share the data which is publicly available.

In RODA, you can use keywords and tags for common types of data such as genomic, satellite imagery and transportation in order to search whatever data that you are looking for. All of this is possible on a simple web interface.

For every dataset, you will discover detail page, usage examples, license information and tutorials or applications that use this data.

By making use of a broad range of compute and data analytics products, you can analyze the open data and build whatever services you want.

While the data you access is available through AWS resources, you need to bear in mind that it is not provided by AWS. This data belongs to different agencies, government organizations, researchers, businesses and individuals.

5. European Union Open Data Portal

You can access whatever open data EU institutions, agencies and other organizations publish on a single platform namely European Union Open Data Portal.

The EU Open Data Portal is home to vital open data pertaining to EU policy domains. These policy domains include economy, employment, science, environment, and education.

Around 70 EU institutions, organizations or departments such as Eurostat, the European Environment Agency, the Joint Research Centre and other European Commission Directorates General and EU Agencies have made their datasets public and allowed access. These datasets have crossed the number of 11700 till date.

The portal enables easy access. You can easily search, explore, link, download and reuse the data through a catalog of common metadata. You can do so for your specific purposes. It could be commercial or non-commercial purposes.

You can search the metadata catalog through an interactive search engine (Data tab) and SPARQL queries (Linked data tab).

By making use of this catalog, you can gain access to the data stored on the different websites of the EU institutions, agencies and organizations.

6. FiveThirtyEight

It is a great site for data-driven journalism and story-telling.

It provides its various sources of data for a variety of sectors such as politics, sports, science, economics etc. You can download the data as well.

When you access the data, you will come across a brief explanation regarding each dataset with respect to its source. You will also get to know what it stands for and how to use it.

In order to render this data user-friendly, it provides datasets in as simple, non-proprietary formats such as CSV files as possible. Needless to say, these formats can be easily accessed and processed by humans as well as machines.

With the help of these datasets, you can create stories and visualizations as per your own requirements and preference.

7. U.S. Census Bureau

U.S. Census Bureau is the biggest statistical agency of the federal government. It stores and provides reliable facts and data regarding people, places, and economy of America.

The Census Bureau considers its noble mission to extend its services as the most reliable provider of quality data.

Whether it is a federal, state, local or tribal government, all of them make use of census data for a variety of purposes. These governments use this data to

determine the location of new housing and public facilities. They also make use of it at the time of examining the demographic characteristics of communities, states, and the USA.

This data is also made use of in planning of transportation systems and roadways. When it comes to deciding quotas and creating police and fire precincts, this data comes in handy. When governments create localized areas of elections, schools, utilities etc, they make use of this data. It is a practice to compile population information once a decade and this data are quite useful in accomplishing the same.

There are various tools such as American Fact Finder, Census Data Explorer and Quick Facts which are useful in case you want to search, customize and visualize data.

For instance, Quick Facts alone contains statistics for all the states, counties, cities and even towns with a population of 5000 or more.

Likewise, American Fact Finder can help you discover popular facts such as population, income etc. It provides information that is frequently requested.

The good thing is that you can search, interact with the data, get to know about popular statistics and see the related charts through Census Data Explorer. Moreover, you can also use visual tool to customize data on an interactive maps experience.

8. Data.gov

Data.gov is the treasure-house of US government's open data. It was only recently that the decision was made to make all government data available for free.

When it was launched, there were only 47. There are now 180,000 datasets.

Why Data.gov is a great resource is because you can find data, tools, and resources that you can deploy for a variety of purposes. You can conduct your research, develop your web and mobile applications and even design data visualizations.

All you need to do is enter keywords in the search box and browse through types, tags, formats, groups, organization types, organizations, and categories. This will facilitate easy access to data or datasets that you need.

Data.gov follows the Project Open Data Schema — a set of requisite fields (Title, Description, Tags, Last Update, Publisher, Contact Name, etc.) for every data set displayed on Data.gov.

9. DBpedia

As you know, Wikipedia is a great source of information. DBpedia aims at getting structured content from the valuable information that Wikipedia created.

With DBpedia, you can semantically search and explore relationships and properties of Wikipedia resource. This includes links to other related datasets as well.

There are around 4.58 million entities in the DBpedia dataset. 4.22 million are classified in ontology, including 1,445,000 persons, 735,000 places, 123,000 music albums, 87,000 films, 19,000 video games, 241,000 organizations, 251,000 species and 6,000 diseases.

There are labels and abstracts for these entities in around 125 languages. There are 25.2 million links to images. There are 29.8 million links to external web pages.

All you need to do in order to use DBpedia is write SPARQL queries against endpoint or by downloading their dumps.

DBpedia has benefitted several enterprises, such as Apple (via Siri), Google (via Freebase and Google Knowledge Graph), and IBM (via Watson), and particularly their respective prestigious projects associated with artificial intelligence.

10. freeCodeCamp Open Data

It is an open source community. Why it matters is because it enables you to code, build pro bono projects after nonprofits and grab a job as a developer.

In order to make this happen, the freeCodeCamp.org community makes available enormous amounts of data every month. They have turned it into open data.

You will find a variety of things in this repository. You can find datasets, analysis of the same and even demos of projects based on the freeCodeCamp data. You can also find links to external projects involving the freeCodeCamp data.

It can help you with a diversity of projects and tasks that you may have in mind. Whether it is web analytics, social media analytics, social network analysis, education analysis, data visualization, data-driven web development or bots, the data offered by this community can extremely useful and effective.

11. Yelp Open Datasets

The Yelp dataset is basically a subset of nothing but our own businesses, reviews and user data for use in personal, educational and academic pursuits.

There are 5,996,996 reviews, 188,593 businesses, 280,991 pictures and 10 metropolitan areas included in Yelp Open Datasets.

You can use them for different purposes. Since they are available as JSON files, you can use them in order to teach students about databases. You can use them to learn NLP or for sample production data while you understand how to design mobile apps.

In this dataset, you will find each file composed of a single object type, one JSON-object per-line.

12. UNICEF Dataset

Since UNICEF concerns itself with a wide variety of critical issues, it has compiled relevant data on education, child labor, child disability, child mortality, maternal mortality, water and sanitation, low birth-weight, antenatal care, pneumonia, malaria, iodine deficiency disorder, female genital mutilation/cutting, and adolescents.

UNICEF's open datasets published on the IATI Registry: http://www.iatiregistry.org/publisher/unicef has been extracted directly

from UNICEF's operating system (VISION) and other data systems, and it reflects inputs made by individual UNICEF offices.
The good thing is that there is a regular update when it comes to these datasets. Every month, the data is updated in order to make it more comprehensive, reliable and accurate.

You can freely and easily access this data. In order to do so, you can download this data in CSV format. You can also preview sample data prior to downloading it.

While anybody can explore and visualize UNICEF's datasets, there are three principal publishers:

UNICEF's AID TRANSPARENCY PORTAL : You can far more easily access the datasets if you use this portal. It also includes details for each country that UNICEF works in.
Publisher d-portal : It is, at the moment, in BETA. With this, portal, you can explore IATI data.
You can search the information related to development activities, budgets etc. You can explore this information country-wise.

Publisher's data platform : On this platform, you can easily access statistics, charts, and metrics on data accessed via the IATI Registry. If you click on the headers, you can also sort many of the tables that you see on the platform. You will also find many of the datasets in the platforms in machine-readable JSON format.

13. Kaggle

Kaggle is great because it promotes the use of different dataset publication formats. However, the better part is that it strongly recommends that the dataset publishers share their data in an accessible, non-proprietary format.

The platform supports open and accessible data formats. It is important not just for access but also for whatever you want to do with this data. Therefore, Kaggle Dataset clearly defines the file formats which are recommended while sharing data.

The unique thing about Kaggle datasets is that it is not just a data repository. Each dataset stands for a community that enables you to discuss data, find out public codes and techniques, and conceptualize your own projects in Kernels.

CSV, JSON, SQLite, Archive, Big Query etc. are files types that Kaggle supports. You can find a variety of resources in order to start working on your open data project.

The best part is that Kaggle allows you to publish and share datasets privately or publicly.

14. LODUM

It is the Open Data initiative of the University of Münster. Under this initiative, it is made possible for anyone to access any public information about the university in machine-readable formats. You can easily access and reuse it as per your needs.

Open data about scientific artifacts and encoded as linked data is made available under this project.

With the help of Linked Data, it is possible to share and use data, ontologies and various metadata standards. It is, in fact, envisaged that it will be the accepted standard for providing metadata, and the data itself on the Web.

The LODUM team has co-initiated LinkedUniversities.org and LinkedScience.org.
You can use SPARQL editor or SPARQL package of R to analyze data.
SPARQL Package enables to connect to a SPARQL endpoint over HTTP, pose a SELECT query or an update query (LOAD, INSERT, DELETE).

15. UCI Machine Learning Repository

It serves as a comprehensive repository of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms.

In this repository, there are, at present, 463 datasets as a service to the machine learning community.

The Center for Machine Learning and Intelligent Systems at the University of California, Irvine hosts and maintains it. David Aha had originally created it as a graduate student at UC Irvine.

Since then, students, educators, and researchers all over the world make use of it as a reliable source of machine learning datasets.

How it works is that each dataset has its distinct webpage which enlists all the known details including any relevant publications that investigate it. You can download these datasets as ASCII files, often the useful CSV format.

The details of datasets are summarized by aspects like attribute types, number of instances, number of attributes and year published that can be sorted and searched.

Open Data Portals and Search Engines:

While there are plenty of datasets published by numerous agencies every year, very few datasets become recognized and established.

The reason why very few such datasets sustain as useful resource is that it is a challenge to develop, manage and provide the data in a way that people and organizations find it useful and easy to use.

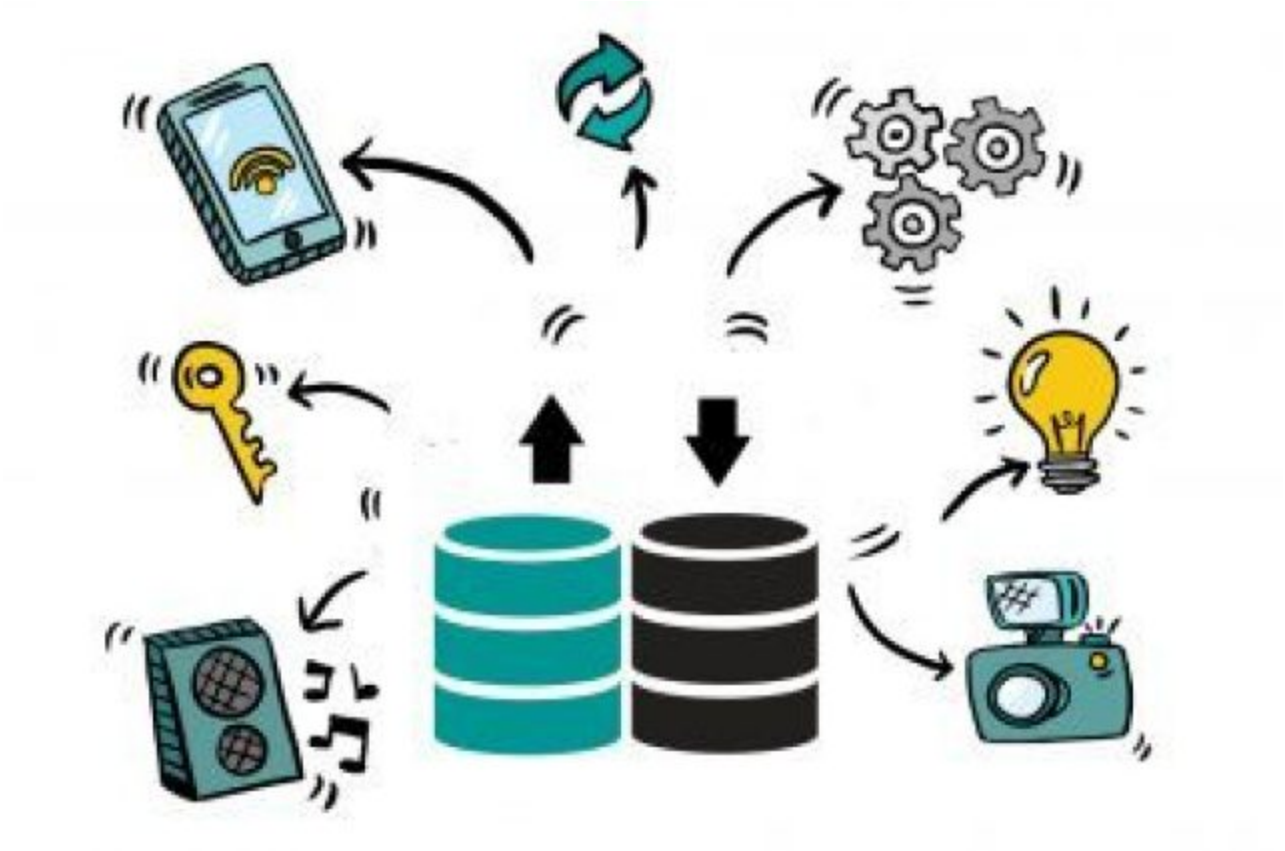
However, please find below a list of other few important open data portals and platforms that permit users to access open data quite easily, study the impact and glean valuable insights.

1. Google dataset search
2. Dataverse
3. Open Data Kit
4. Ckan
5. Open Data Monitor
6. Plenar.io
7. Open Data Impact Map

Conclusion

Open data is the order of the day. The world has gradually started moving towards open systems and open data is rightly in sync with that

Open Source Databases that Work Best for IoT



The Internet of Things (IoT) generates vast amounts of data, including streaming data, time series data, RFID data, sensory data, etc. The efficient management of this data demands the use of a database. The very nature of IoT data requires a different type of database. Here are some databases that give very good results when used in conjunction with IoT.

The Internet of Things (IoT) can be regarded as a network in which various things are connected to each other through a common platform. Just visualise a scenario in which every device at home and the workplace is connected, and a world where the air-conditioning in a room automatically lowers its temperature when the outside temperature rises up, when the number of people in any public gathering is easily known, and when one's health parameters can be monitored on a daily basis. This is the possible impact of the Internet of Things.

The current state of the Internet of Things is very fragmented. There are different companies and organisations that are building their own platforms for either their customers or their individual needs. But a common platform on which all the devices, irrespective of their company, can be connected with each other via a user friendly interface, is still missing.

IoT devices are estimated to number in the trillions in the coming five years.

Is a database necessary for IoT?

The Internet of Things creates many tedious challenges, especially in the field of database management systems, like integrating tons of voluminous data in real-time, processing events as they stream and dealing with the security of data. For instance, IoT based traffic sensors applied in smart cities would produce huge amounts of data on traffic in real-time.

Databases have a very important role to play in handling IoT data adequately. Therefore, along with a proper platform, the right database is equally important. As IoT operates across a diverse environment in the world, it becomes very challenging to choose an adequate database.

The factors that should be considered before choosing a database for IoT applications are:

1) Size, scale and indexing

2) Effectiveness while handling a huge amount of data

3) User-friendly schema

4) Portability

5) Query languages

6) Process modelling and transactions

7) Heterogeneity and integration

8) Time series aggregation

9) Archiving

10) Security and cost

The types of data in the Internet of Things are:

1) *RFID*: Radio frequency identification

2) Addresses/unique identifiers

3) Descriptive data for processes, systems and objects

4) Pervasive environmental data and positional data

5) *Sensor data:* Multi-dimensional time series data

6) Historical data

7) *Physics models:* Models that are templates for reality

8) State of actuators and command data for control

Databases suited for the Internet of Things

InfluxDB: InfluxDB was first released in 2013, and is one of the recent databases. The Go programming language was used in developing this database, which is totally based on LevelDB, a key-value database. InfluxDB is a time series database,

which is used to optimise and handle time series data. Time series data was first released by Kdb in 2000, but InfluxDB became popular with the rise in the Internet of Things as it gave movement to NoSQL, NewSQL and a vast amount of increasing data.

The advantages of using InfluxDB for IoT data include:

1) Allows indexing of series

2) It has an SQL-like query language

3) It also provides the built-in linear interpolation for missing data

4) It supports automatic data down sampling

5) Supports continuous queries to compute aggregates

CrateDB: CrateDB is a distributed SQL database management system. Being open source and written in Java, it includes components from Facebook Presto, Apache Lucene, Elasticsearch and Netty—thus it is designed for high scalability. CrateDB was made for putting IoT data to work. From the industrial Internet and connected cars to wearables, CrateDB is the database of choice for innovators of new IoT solutions.

The advantages of using CrateDB for IoT data include:

1) *Millions of data points per second:* Fast, linearly scalable data ingestion

2) *Real-time queries:* Columnar indices and field caches provide in-memory SQL performance

3) *Dynamic schema:* Add and query new sensor data structures on-the-fly

4) *IoT analytics:* Fast, robust time series, AI, geospatial, text search, joins, aggregations

5) *Always on:* Built-in data replication and cluster rebalancing ensure non-stop performance

6) *ANSI SQL:* No lock-in, and easy for any developer to use and integrate

7) *Built-in MQTT broker:* Direct device-to-database integration

8) *IoT ecosystem:* Works with Kafka, Grafana, NodeRED, and other popular IoT stack software

9) Runs anywhere for efficient processing at the edge or in the cloud

MongoDB: MongoDB is a free and open source cross-platform document-oriented database program. It is categorised as a NoSQL database program. JSON-like documents with schemas are used by MongoDB. It is preferred by organisations for IoT, as it lets them store data from any context, which can be analysed in real-time, and also to change the schema as they go along.

The advantages of using MongoDB for IoT data include:

1) Highly powerful database

2) Document-oriented

3) Has uses for general purposes

4) Being a NoSQL database, it uses JSON-like documents with schemas

RethinkDB: In the open source database list, RethinkDB stands at the top. It is a scalable JSON database for the real-time Web, which is built from the ground up.

RethinkDB introduces an exciting new access model by transposing the traditional database architecture. It can continuously push updated query results to applications in real-time, when a command is given to it by the developer. This is a feature the developers call changefeeds. RethinkDB serves as a database, real-time repository and message broker of the system state, which is allowed by changefeed. Its real-time push architecture dramatically reduces the time and effort necessary to build scalable real-time apps.

The advantages of using RethinkDB for IoT sensor data include:

1) RethinkDB has an adaptable query language for examining APIs, which is very easy to set up and learn.

2) Commands are automatically shifted to a new server if any primary server fails.

3) Plug-and-play function of nodes in real-time, without any downtime for even a single second, helps in the easy addition of nodes.

4) Offers asynchronous queries via Eventmachine in Ruby and Tornado, which gives an asynchronous application programming interface.

5) It offers SSL access just to have secured access to RethinkDB via public Internet.

6) Floor, ceil and round are various mathematical operators that are offered by RethinkDB.

SQLite: SQLite Database Engine is a process library that provides a serverless (self-contained) transactional SQL database engine. It has had a major impact on game and mobile application development due to its portability and small footprint.

SQLite works appropriately with the devices that do not require any human support, as the database requires no administrative permissions. It is a good fit for use in cell phones, set-top boxes, televisions, game consoles, cameras,

watches, kitchen appliances, thermostats, automobiles, machine tools, air planes, remote sensors, drones, medical devices and robots, as well as in IoT.

Client/server database engines are designed to live inside a data centre at the core of the network. SQLite works there too, but SQLite also thrives at the edge of the network, fending for itself while providing fast and reliable data services to applications that would otherwise have dodgy connectivity.

The advantages of using SQLite for IoT data include:

1) Offers a small memory footprint

2) It is authentic

3) No setting up required prior to use

4) Has no dependencies

Apache Cassandra: Apache Cassandra is a free and open source distributed NoSQL database management system, which was initially released in 2008. It was intended to handle huge amounts of data through many commodity servers, providing high availability with no single point of failure.

In IoT, the generation, tracking and sharing of data through a variety of networks is carried out on an immense scale due to the massive number of connected devices. Cassandra is excellent at utilising lots of time series data that comes directly from devices, users, sensors, and similar mechanisms that subsist in diverse geographic locations.

The advantages of using Apache Cassandra for IoT

data include:

1) Fault tolerant

2) Demonstrates high performance

3) *Decentralised:* Every node in the cluster is identical

4) Scalable

5) Durable

6) *Ensures you're in control:* Each update has a choice of synchronous and asynchronous replication

7) *Elastic:* Both read and write execute in real-time, thus there is no downtime for any application

8) *Professionally supported:* It reinforces contracts and services that are available from third parties.