

Emerging Trends in Database and its Applications in MNC's

Anushri Gupta*

Department of Computer Science and Engineering, Sagar Institute of Research and Technology, Bhopal, Madhya Pradesh, India

Email: anushrigupta4302@gmail.com

ORCID iD: <https://orcid.org/0009-0009-5623-6438>

*Corresponding Author

Ritu Shrivastava

Department of Computer Engineering, Sagar Institute of Research and Technology, Bhopal, Madhya Pradesh, India

Email: ritushrivastava08@gmail.com

ORCID iD: <https://orcid.org/0009-0004-1342-722X>

Jay Kumar Jain

Department of Computer Science and Engineering, Sagar Institute of Research and Technology, Bhopal, Madhya Pradesh, India

Email: jayjain.research@gmail.com

ORCID iD: <https://orcid.org/0000-0002-9590-0006>

Received: 02 February, 2023; Revised: 08 March, 2023; Accepted: 18 April, 2023; Published: 08 August, 2023

Abstract: When information is set for easy retrieval, access, updation, and management, then it is known as Database. Its major purpose is storing aggregations of records and data files that contain information regarding transactions, customer details, product details and financials. The paper deals with a briefing on DBMS which is a collection of inter-related data and set of programs to retrieve and access the data which focuses on :- a) exchange of data, b) retrieve the data, c) storage of data, and d) manipulation on data. The paper also gives a glance on elements of DBMS also describes the architecture which is consist of: a) internal level, b) conceptual level, and c) external level. The paper also mentions some of the uses of databases and will try to give an idea about the applications of Database in different MNC's and projects in Data ware.

Index Terms: Database and it's types, Components, Big Data and it's applications, DBMS and it's elements (about DBA and Architecture), Uses and Applications in MNC's and projects.

1. Introduction

Databases were created in 1960's, while Relational Databases were first developed in 1970's and Object-oriented Databases were first developed in 1980's. Earlier databases were network models which consisted of records related to primary as well as secondary records.

The basic objective behind using Database is to eliminate redundancy, build a design of data, and ensure the usability, extensibility, data integrity and performance. The limitation faced in most of the databases are : high complexity, high cost, database failure, large amount of storage required, upgradation costs and cost for data conversion[1].

Database Management System provides a facility for the end users for creating, updating and deleting the data.

2. Types of Database

- **Musculoskeletal disorders (MSDS) Relational Database:** This approach is reorganized in form of tables and data can be accessed through rows and columns. Columns define at least one (or more) category whereas rows have data instance for the categories defined in the columns. For application program interfaces and user, relational database uses SQL. Relational Database Management System (RDBMS) is used for managing,

storing and retrieving the queries with ACID properties (atomicity, consistency, isolation and durability).

- **Distributed Database:** This database is used for storing records/files in physical locations. These are homogeneous and have same hardware and run on same operating system.
- **Cloud:** These databases are used to be built on private, public and hybrid cloud for virtualization environment and work with applications deployed as software as a service.
- **NoSQL:** These database deal with large collection of distributed data and can address the big data performance issues. These are also known as non-relational databases[2].
- **Object-oriented database:** These hold data which is created by the object oriented programming languages and focuses on object organization.

Components of Database

- ✓ **Data**
- ✓ **Hardware**
- ✓ **Software**
- ✓ **Procedure**
- ✓ **Data Access Language**

3. Some emerging Database technologies with features

After reading the literature Database can be changed according to the company needs and its requirements. We need database to secure and protect data from an unauthorized users. [3,4].

Table 1. Databases and their features [14]

DATABASE	FEATURES
MongoDB	• Company-MongoDB Inc (with Server side Public License), • It uses Ad-hoc queries, • It provides the facility of indexing, • It allows replications, • It also ensures high file storage in the system.
DB2	• Company - IBM • It provides facilities that supports multiple data types & languages, • It gives feature of advanced protection of data throughout, • It has the industry leading performance, and • It mainly deals with the continuous availability of data in the database.
MySQL	• Open source RDBMS • It gives high and best performance, • The data stored is compatible & portable, • Provides ease of database management, • This database is reliable and fast.
NoSQL	• Open source RDBMS • It provides the facility of the multiple data model compatibility, • It enhance the scalability & availability of data, • It gives the global data distribution, • This database has the minimal downtime.
PostgreSQL	• Open source object relational database, • Current version used is 9.6.2, • Provides high data security, • Data retrieval is faster in this database, • Data sharing is done through dashboards.
Oracle	• Company-Oracle, • Provide secure hybrid cloud environment, • Executes quick backup & recovery, • Includes the concept of flashback technology, and • Provides superior scalability.

Amazon RDS	• Offered by Amazon Web Services (AWS), • Supports database engines which are used for storage and organization of data, • Supports activities related to relational database maintenance, backup and recovery, data migration and patching, • Gives high availability, security and compatibility of data required, • Concentrates on applications which give quick responses.
Neo4j	• Developed by – Neo4j.Inc • Latest version used is 3.1.10, • It is also known as Graph DBMS as it stores data in the form of graphs, • Has large capacity server, • Used on Windows and Unix\ Linux operating system.
OrientDB	• Company - OrientDB Ltd • Latest version used is 2.2.17, • It is a graphical database, • Widely used across big data market, • It has real-time web-based application.
MS SQL Server 2022	• Company-Microsoft, • It is used for the analytics purpose as well , • It ensures the availability of the query store & intelligent query processing, • It focuses on the data management, • It has high security, availability and performance of data.

4. Introduction to Big Data

Big data refers to large and complex data of datasets that are difficult to store, process and manage. Major challenges include- analysis, data capture, searching data, data transfer, data privacy and data visualization.

4.1. 4 V's of Big data

- ✓ Volume
- ✓ Velocity
- ✓ Value
- ✓ Variability

In other words, Big Data is high-volume, high-velocity and high-variety information set which demands cost-effective and innovative techniques of processing data for decision making purpose [5].

4.2. Some analysis tools of Big Data -

- **Apache Hadoop** – It is open source framework which can run in cloud and has less hardware demand and consists of features like – HDFS, MapReduce, YARN and Hadoop libraries.
- **HDFS (Hadoop Distributed File system)** is the storage unit where data is stored to run map reduce jobs.
- **MapReduce** is used for computation of logics which are based on Hadoop for providing fast data processing and so it is also known as the fundamental algorithm of Hadoop.
- **Daemons of Hadoop** – **a) Namenode** (runs on HDFS master node), **b) Datanode** (runs on HDFS slave node), **c) Secondary Namenode** (backup for namenode in case of failure of namenode), **d) JobTracker** (communicates with namenode and locates the TaskTracker near the data), and **e) TaskTracker** (accepts the tasks assigned by JobTracker and transfers to the execution slots which are available on the system).[6,7]
- **MongoDB** – It is also an open source NoSQL database which is cross-platform compatible with different programming languages and is used in cloud computing. It has features of: cloud-native deployment, flexibility of configuration, can store any type of data, data partitioning, etc.
- **Apache Spark** – It provides in-memory data processing facilities which is faster than disk processing done by MapReduce. [8]

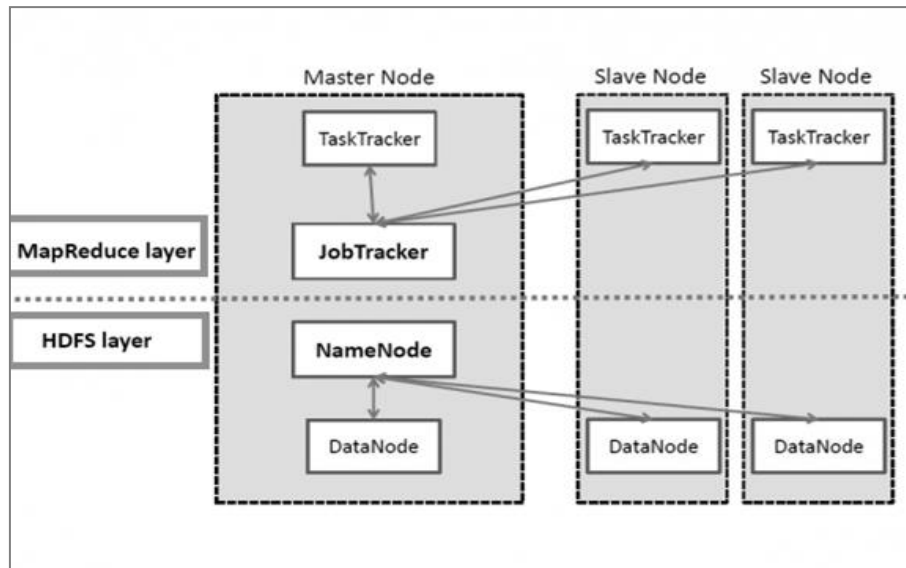


Fig. 1. Architecture of Hadoop (HDFS and MapReduce)

4.3. Applications of Big Data-

- **In Healthcare** – which includes patient's data, EHR data, fitness trackers, transaction data etc.
- **In Manufacturing** – which includes product quality and defect tracking, supply planning, manufacturing process defect tracking, increasing efficiency of products etc.
- **In Media & Entertainment** – predicting demand of audience, scheduling optimization, Ad targeting, product development etc.
- **In Governance** – cyber security, crime prediction and prevention, pharmaceutical drug evaluation, some scientific research, weather forecasting etc.[9]

5. Database Management System and it's elements

A DBMS provides facility to enable the users for creating and managing the database, read and update the data, and also assists the users with auditing and logging functionalities. It also provides some physical as well as logical independence from the data in the database.

The challenges for database are—

Data consistency, data redundancy, data independence, data security, data integrity, and data isolation.

5.1 Elements Of Dbms-

- ✓ Data
- ✓ Relation
- ✓ Schema
- ✓ Constraints

1. **Data** – It refers to items stored in a database in the form of database tables (organised into rows and columns).
2. **Relation** – It refers to the association between two or more data sets. It is of 4 types which are:
 - a) One-to-one relationship
 - b) One-to-many relationship
 - c) Many-to-one relationship
 - d) Many-to-many relationship
3. **Schema** – It is the overall structure and organization of data in the database represented in some formal language.
4. **Constraints** – For ensuring the accuracy, reliability in a database and to limit the data entry into the table, constraints are used. In case of violation between the data and constraint, action does not take place. [10]

5.2 Database Administrator (DBA) –

It is responsible for performing all the major activities in database for maintaining database environment.

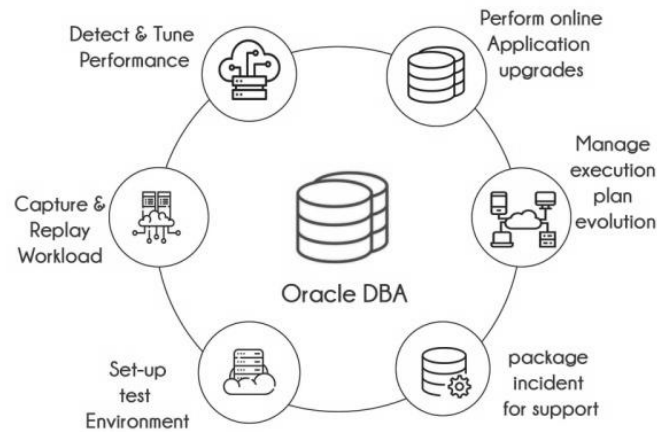


Fig. 2. Features of DBA

Its major benefits are:-

- Speeds the decision making process,
- Enhances the data security as well as integrity, and
- Helps in predicting the behavior of customers.[11]

5.3 Architecture of DBMS:-

The architecture of DBMS is known as three schema architecture or three-level architecture.

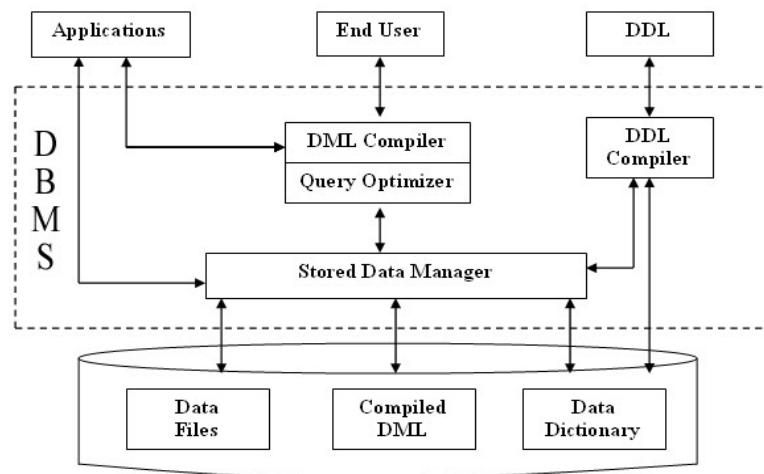


Fig. 3. Architecture of DBMS

The architecture of DBMS consists of:

- ✓ External Level
- ✓ Conceptual Level
- ✓ Internal Level

External Level – This level is also known as View Level. Its consist of number of external schemas as well as user views. It describes only the part of database that the user is interested in and hides the rest of the database.

Conceptual Level – It stores information about what the data is to be stored in the database. It describes the structure of entire database concentrating on entities, data types, operations, relations and constraints and hiding physical storage structures.

Internal Level – It describes the physical storage structure of entire database and gives information about how the data is actually stored in the database.[12]

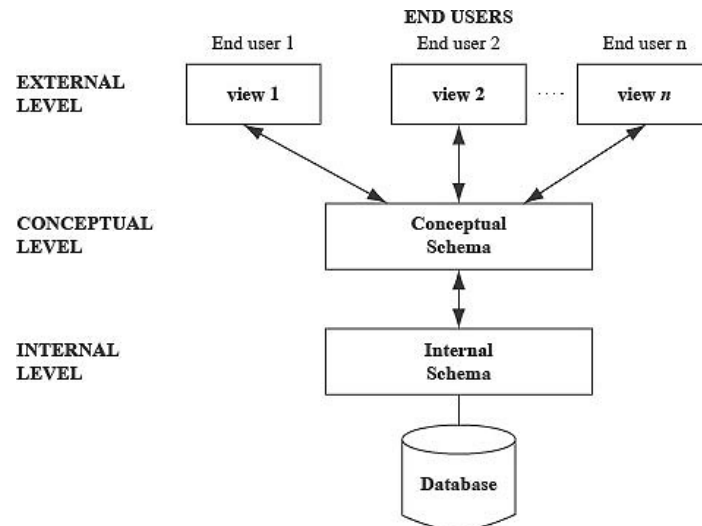


Fig. 4. Three level architecture of database.

6. Uses of Database Management System

Databases are mainly used to sort, access and maintain any data or information that can be analyzed, or in other words, it is the organized collection of data/information.

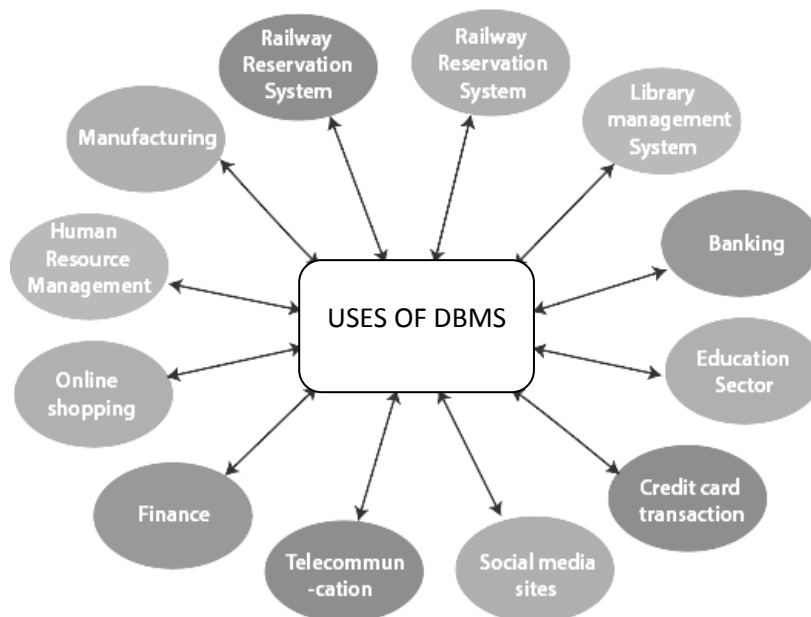


Fig. 5. Uses of DBMS

Some daily use examples of databases are:

- **On demand online video streaming:** Online streaming videos make use of databases for keeping track of different shows, available movies and view preferences to provide better watch recommendations example Apache Cassandra (distributed database).
- **Social gaming:** Database keeps record player information and serves other players on-demand. Example– Game of Thrones Ascent which is free role-playing game.
- **Personal cloud storage:** Google Drive, Microsoft OneDrive, iCloud etc. are examples of personal cloud storage devices that stores complex data models and powerful data warehouses to ensure data security.
- **Finances and E-commerce:** The tracking of enormous amounts of information from global daily transactions requires powerful and secure databases and use different models that analyse data to predict future activities, and also organise the products, customer information, pricing and purchasing history.

- **Government organizations:** Many different Government organizations worldwide contains large amounts of data including researches, defence, and legislation.
- **Social media:** According to the statistics, 500+ terabytes of new data gets ingested into the databases of social media site Facebook per day.
- **Healthcare:** Different healthcare sectors and organizations store and analyse large and extensive amounts of patient data with complex data structures and security protecting sensitive data in databases that is easy to access and manage.
- **Weather:** Different weather organizations use prediction models, the data is stored and analysed within the databases that is accessible and deliverable on TVs and other apps, including Microsoft SQL Server. [13]

7. Applications of Database in MNC's

- ✓ The input data is entered in the form of CSV, TEXT, JSON, PARQUET.
- ✓ The input files are then placed at different FTP Sites by the users.
- ✓ The stages involved are as follows:-
 - The File Watcher Job runs at defined interval to find whether the file has arrived. (The automation tools include Cyber-automation, Autosys, etc)
 - (File Watcher job will have code written in Python/UNIX to read the input file as soon as it comes)
- ✓ After File Watcher job gets to know about that the input file has arrived, it will call another job which will process the input file.
- ✓ File Processing Job calls UNIX shell script to load the input file to Oracle Data ware House.
- ✓ UNIX shell script calls Oracle PLSQL package. In package, procedures will be used to load data to Data ware House.
- ✓ Through Oracle packages only, data will be cleaned and loaded into final target table in format required by user.
- ✓ Filtered and transformed data from tables will be used by the BI tools to make reports for the users.
- ✓ After all the steps completed, mail is sent by the job users to review the data. (For the mail also job is used which sends mail at the end)
- ✓ In case of any job failure during loading, mail will be sent to the developers or to the supporters for fixing the issue. [15]

8. Conclusion

In conclusion, the database gives the centralized facility for quick sharing of data among different users and easy modification of data along with facility of storing and organizing data and so it is used by companies with massive data. Therefore, DBMS in other words is a systematic operational system consisting of different tools used for making data, saving data, data manipulation and handling data.

In this paper, the briefing about the architecture of DBMS, and is mainly focusing on the uses in gaming, governance, healthcare sector, forecasting, cloud computing. The paper also gives briefing on the application of databases in different MNC's especially in the Dataware housing process on different FTP Sites using the File Watcher Jobs which gives File Processing calls and UNIX shell calls. Database is thus necessary to set up for easy access of data, management of data and also for updating the data.

References

- [1] Codd, E. F. (1970). A Relational Model of Data for Large Shared Data Banks Retrieved on June 12, 2012, from <http://www.seas.upenn.edu/~zives/03f/cis550/codd.pdf>
- [2] Sumathi, S., Esakkirajan, S., (2007). Fundamentals of Relational Database Management Systems. (pp. 96- 97).Springer-Verlag Berlin Heidelberg.
- [3] Chauhan, Dipti, Jay Kumar Jain, and Sanjay Sharma. "An end-to-end header compression for multihop IPv6 tunnels with varying bandwidth." *2016 Fifth international conference on eco-friendly computing and communication systems (ICECCS)*. IEEE, 2016.
- [4] Cocea, M., and Weibelzahl, S (2009), "Log file analysis for disengagement detection in e-learning environments", Springer Journal User Modeling and User Adapted Interaction. Vol.19, No.4, pp.341-385. DOI: 10.1007/s 11257-009-9065-5
- [5] M. D. Assuno, R. N. Calheiros, S. Bianchi, M. a. S. Netto and R. Buyya, Big data computing and clouds: Trends and future directions, Journal of Parallel and Distributed Computing, 79 (2015), pp.3-15.
- [6] K. Kambatla, G. Kollias, V. Kumar and A. Gram, Trends in big data analytics, Journal of Parallel and Distributed Computing, 74(7) (2014), pp.2561-2573.
- [7] D. P. Acharjya, S. Dehuri and S. Sanyal Computational Intelligence for Big Data Analysis, Springer International Publishing AG, Switzerland, USA, ISBN 978-3-319-16597-4, 2015.

- [8] G. Ingersoll, Introducing apache mahout: Scalable, commercial friendly machine learning for building intelligent applications, White Paper, IBM Developer Works, (2009), pp. 1-18.
- [9] Abraham Silberschatz, Henry F. Korth and S. Sudarshan, Database System Concepts, McGraw-Hill Education (Asia), Fifth Edition, 2006.
- [10] Hershey, William; Easthope, Carol (1972). A set theoretic data structure and retrieval language. Spring Joint Computer Conference, May 1972. ACM SIGIR Forum. Vol. 7, no. 4. pp. 45–55. doi:10.1145/1095495.1095500
- [11] Jain, Jay Kumar, Devendra Kumar Jain, and Anuradha Gupta. "Performance analysis of node-disjoint multipath routing for mobile ad-hoc networks based on QOS." *International Journal of Computer Science and Information Technologies* 3.5 (2012): 5000-5004.
- [12] King, E. (2011). The Growth and Expanding Application of In-Memory Databases Retrieved from <http://www.loyola.edu/lattanze/workingpapers/WP0611-107.pdf>
- [13] Jain, Jay Kumar, and Sanjay Sharma. "Performance Evaluation of Hybrid Multipath Progressive Routing Protocol for MANETs." *International Journal of Computer Applications* 71.18 (2013)
- [14] Gupta, D., Jindal, R., Dutta Borah, M (2011), "A Knowledge Discovery based Decision Technique in Engineering Education Planning" in Proc. Int. Conf. On Emerging Trends & Technologies in Data management. Institute of Management Technology Ghaziabad, pp. 94-102.
- [15] Beynon-Davies, Paul (2003). Database Systems (3rd ed.). Palgrave Macmillan. ISBN 978-1403916013.

Authors' Profiles



Anushri Gupta is currently working as a pre final year student in the Department of Computer Science and Engineering, Sagar Institute of Research and Technology, Bhopal, India. She has published 1 research paper in International Journal of Scientific Research in Engineering and Management (IJSREM) Her research interest in Python, Internet of Things and DBMS.



Dr. Ritu Shrivastava did her Btech in Computer Science and Engineering from SATI Vidisha in 1992 and completed her Phd. In Computer Science and Engg in the year 2014. Her area of specialization includes Web Engineering, Data Mining, Big Data and Cloud Computing. She has over 25 years of teaching experience. The experience includes 10 years in the Computer Science Department of "Maulana Azad National Institute of Technology (MANIT)" at Bhopal. MANIT which is among the seven National Institutes of Technology in India, commanding a premium from best of the students and industries across India. She has also worked with Rai University, New Delhi and Amity Institute of Information Technology, Noida. She worked at Babu Banarsi Das National Institute of Tech.& Management from March'06 to February'09. For the last 14

years She is Working in Sagar Group of Institutions in various capacities. During the course of her 25 Years she has published many papers in National and International Journals. She has organized many conferences and FDPs in the Institute. She is a Life member of CSI, ISTE and also member of many Professional bodies. She is also on the Editorial board of many reputed Journals. She has guided many UG and PG Students. She has published 1 patent and filed 2 patents. She has published 2 copyrights and filed 2 copyrights. She has guided over 75 UG and PG students.



Dr. Jay Kumar Jain is is currently working as an Associate Professor in CSE Department & MCA (Head), Sagar Institute of Research & Technology, Bhopal. He did his Ph. D. from Maulana Azad National Institute of Technology, Bhopal in 2015. He was awarded with Research Fellowship by MHRD for completing his Ph.D. He has research as well as teaching experience of about 16 years. He has published around 50+ research papers and book chapters in SCI/SCOPUS/Referred International/National Journal and Conferences. He has also granted 2 International Patents and published 2 Indian patents and also registered 2 copyrights in IPR, India. He has been a reviewer in many international journals/conferences including Elsevier, IEEE Access, and Springer. He has lifetime membership of various professional societies such as CSI, Franklin, IDES, IAENG, SDIWC, and many more. He has also working as an innovation Coordinator and Ambassador in Institution's Innovation Council (IIC), SIRT Bhopal, under Ministry of Education (MoE), Govt. of India. His research interests include Wireless Sensor Networks, Internet of Things, and Mobile Ad hoc Networks.

How to cite this paper: Anushri Gupta, Ritu Shrivastava, Jay Kumar Jain, "Emerging Trends in Database and its Applications in MNC's", International Journal of Education and Management Engineering (IJEME), Vol.13, No.4, pp. 33-41, 2023. DOI:10.5815/ijeme.2023.04.05