

Next-Gen Big Data Management: The Role of In-Memory Databases

Dr. Ravi Prakash¹, Devika Rani Roy²,

¹ Associate Professor, Department of Computer Engineering,
K. C. College of Engineering & Management Studies & Research, Thane.

² Assistant Professor, Department of Information Technology,
K. C. College of Engineering & Management Studies & Research, Thane.

Abstract: Digital communication generates large amounts of data. This is a great opportunity for companies that work with big data. However, the more data a company has to work with, the greater the challenge becomes to recognize connectivity and patterns. Information technology solutions and systems are ever more in demand to support companies in evaluating the huge amounts of information they receive. Data analysis using traditional databases is no longer sufficient to store, retrieve, and process extremely large collections of data. When classic databases reach their limits, in-memory databases can be of use.

Keywords: Big Data, Digital communication, databases, Data Analysis, In-Memory, information.

1.0 Introduction

Large amounts of data are stored in In-memory databases and it also provide a large range of analysis results. When storing data with an in-memory database, a distinction is between column-oriented and row-oriented data storage is there, but some database systems use both methods of data storage i.e column-oriented and row-oriented data storage. In one row, Row-oriented database are arranged and the collected data records kept together .

For example, if the values “student name, rollno , and subject” are stored, the data would be arranged as follows: student name 1, rollno 1, subjec 1, student name 2, rollno 2, subjec 2 ,student name 3, rollno 3 subjec 3. In a column-based storage, the data is to be assigned in corresponding categories like in this way: student name1,student name2,student name3,rollno1,rollno2,rollno3, subject1, subject2, subject3.

The format of the column-based data storage is called the column format. By keeping data with the same values together, the system reduces the amount of data available. Storage space and transfer times are reduced. The functionality of the memory database has also improved over time, and only certain columns need analysis, not all. This form of data analysis is called columnar projection.[1]

2.0 Technology for big data storage

The concept of in-memory databases is nothing new. The foundations of database technologies were developed in mid-1980s. However, the Information technology systems back then did not have the required processing capacity, so that earlier concepts of in-memory databases were not are used. Modern computer architectures use the concept of data warehousing, 64-bit technology, and multi-core processors finally made it possible for in-memory databases and its use.

Data inside the memory is usually for data storage. These database systems collect and compress data from a variety of sources, store it for a long time, and then prepare it for analysis. With 64-bit technology, it is possible to increase the capacity of large memory up to a terabyte range. As a result, memory flows have increased in size.[1],[2]

With multi-core processors, most processor cores operate on a single chip, resulting in better performance and higher data performance. Data operation reflects the volume of the transmitted network data.

2.1 Steps of an In-Memory Database

Recurring, identical processes occur during the running of in-memory databases. An in-memory database back up the data in the following way:

1. **Start the database:** when the database is started, the system loads the entire dataset from the hard disk into the working memory. It means that no data has to be loaded while the database is running.
2. **Readjusting data:** the database will review and adjust data frequently, so if data changes it remains up to date.
3. **Transaction log and its backup:** current changes are recorded in transaction logs. If an error occurs, the database can be restored to the time before the error occurred. This process is called "rollforward".
4. **Data processing:** data is processed according to the ACID principle (atomicity, consistency, isolation, and durability), as it is in traditional databases. The acronym ACID describes the exact properties of processes in database management systems.
5. **Database replication:** this step continuously copies data from the database to a computer or server as a backup.[3]

2.2 In-memory analytics

It is an enterprise architecture framework solution that is used to enhance business intelligence reporting by querying data from system memory versus the traditional hard disk drive medium. According to this approach, inquiry time is reduced in an effort to facilitate efficient business decisions.[2]

With the development of business intelligence and random access memory hardware technology, more business intelligence platforms are available and affordable, including in-memory analytics tools that are used to facilitate enterprise decision-making, even for small businesses. Traditional BI Online analytical processing incorporates dedicated, heavy resources for programming or building data structures through unusual schemas.[2],[3],[4]

In-memory analytics eliminate the overhead of storing data aggregate tables or indexing pre-aggregated data cubes, resulting in extremely fast query responses. In short, faster data retrieval rates indicate faster processing and informed decision making.[4]

3.0 Big Data Analytics Strategy

A well-defined, integrated and comprehensive strategy contributes to and supports valuable data-driven decision making in an organization. Here, I listed the most followed step.

Step 1: Evaluation: An assessment, which is already tied to business objectives, involves key stakeholders, forming a team of members with the right skill sets and evaluating policies, people, process and technology and data assets Is required. In this process, evaluation clients may be required, if necessary.

Step 2: Priority: After evaluation, one needs to get use cases, prioritize those using big data predictive analytics, prescriptive analytics and development analytics. One can also use tools like priority matrix and filter the use cases with the help of feedback and input from the main user.

Step 3: Roadmap: In this phase we can create a timed roadmap and publish it to everyone. It needs to include all the details about the complexities, funding, inherent benefits of use cases, and mapped projects.

Step 4: Change Management: Such a change requires managing data availability, integrity, security, and usability. A successful change management program, using any existing data governance, encourages activities and members on a continuous monitoring basis.

Step 5: The Right Skill Set: Identifying the right skill set is critical to the success of the organization amidst current trends in the industry. In this phase, there is a need to bring educational programs to follow the right leaders and educate key stakeholders.

Step 6: Reliability, scalability and security: The right approach and effective Big Data analytics strategy make the analytics process reliable, with the powerful use of explanatory models incorporating data science principles. A big data analytics strategy needs to incorporate security aspects right from the beginning for a strong and tightly integrated analytics pipeline.[5]

4.0 Advantage and disadvantage of in-memory databases

The advantage of in-memory databases is that access to data is much faster than traditional databases. The disadvantage of in-memory databases is that it is not possible to store data permanently on RAM. Here is a comparison of the advantages and disadvantages of in-memory databases.

4.1 Advantage of in-memory databases

The biggest advantage of using in-memory databases is the high access speed arising from the use of RAM. It also leads to a quick data analysis. However, it is not only the reduced fetch time that optimizes data analysis. In-memory DBs make it possible to evaluate structured and unstructured data from any system. Now, companies and software solutions are expected to face the challenge of storing and processing large-size unstructured data, such as text, images, or audio and video files.[6]

Using a distributed data infrastructure, unstructured data can be stored in an in-memory database, in which multiple processing units (computers, processors, etc.) operate on a common task in parallel and distribute it to different server groups. This results in a higher storage capacity, faster processing, and better transfer speed

of unstructured data.

4.2 Disadvantage of In-Memory Databases

The use of RAM means faster access to one side, but it also has a significant disadvantage: stored data is only temporary. If the computer system crashes, then all the data stored temporarily will be lost. To add to this, the following methods are established:

1. Snapshot files: are specific moments, such as at regular intervals or before the system is shut down, when the current version of the database is saved. An important criticism of this is that all the data added after the last so-called snapshot will be lost in case of a crash - depending on how long each interval is, it can be a lot of data.

2. Transaction Protocol Security: Noting changes in transaction logs is a unified process that is used as a means of protection. Used in combination with a snapshot process, the transaction protocol can help restore the system after a crash.

3. Replication: In-memory databases already have the function of storing an exact copy of the database on a traditional hard disk. In the event of a failure, the stored database can be accessed.

4. Non-volatile RAM: A non-volatile RAM is able to keep files available for recovery even after the system is restarted.[5],[6]

5.0 List of In-Memory Databases

- 1. H2 Database:** H2 is an open source database written in Java that supports standard SQL for both embedded and standalone databases. It is very fast and contains only 1.5 MB of Java archive files.
- 2. HSQLDB (HyperSQL Database):** HSQLDB is an open source project, also written in Java, which represents a relational database. It follows structured query language and Java database connectivity standards. It also supports SQL features such as stored procedures and triggers. It can be used in in-memory mode, or it can be configured to use disk storage.
- 3. Apache Derby Database:** Apache Derby is another open source project that has a relational database management system created by the Apache Software Foundation.
- 4. Derby:** It is based on SQL and JDBC standards and is primarily used as an embedded

database, but can also be run in client-server mode using the Derby Network Server Framework.

5. **SQLite:** This is a SQL database that runs only in embedded mode, either in memory or saved as a file. It is written in C language but can also be used with Java.
6. **In-Memory Database in Spring Boot:** Spring Boot makes it particularly easy to use in-memory databases because it can automatically create configurations for H2, HyperSQL databases, and Derby.
7. **HyperSQL Database:** It uses an in-memory structure for faster operations against the DB server. It uses disk persistence according to user flexibility with reliable crash recovery. It is also suitable for business intelligence, ETL and other applications that process large data sets. It has a wide range of enterprise deployment options, such as XA transactions, connection pooling data sources, and remote authentication. It is written in the Java programming language and runs in a Java virtual machine. It supports the JDBC interface for database access.[4],[5],[6]

6.0 References

- [1.] Omae Malack Oteri, Mobile Subscription, Penetration and Coverage Trends in Kenya's Telecommunication Sector.
- [2.] Hoda A. Abdel hafez Mining Big Data in Telecommunications Industry: Challenges, Techniques, and Revenue Opportunity.
- [3.] A hybrid fuzzy-based Personalized Recommender System for Telecom Products/Services.
- [4.] Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification Tina R. Patil, Mrs. S. S. Sherekar Sant Gadgebaba Amravati University.
- [5.] Big Data & Advanced Analytics in Telecom: A Multi-Billion-Dollar Revenue Opportunity.
- [6.] www.omnisci.com/in-memory-databa