

Factors to be considered while choosing the right database cont'd

8. Disaster Recovery

Disaster Recovery is another important aspect while choosing the database. Disaster recovery plans must be properly reviewed to ensure that the chosen database engine meets the requirements. If database takes too long to backup and restore, it may not be a suitable one. And if backup is not an option in the Database, then manual backup & restore with bunch of flat files is required. Disaster recovery at a higher level is not required if our application is not so critical.

9. Data shape

RDBMS and typical SQL databases stores data in tables in form of rows and columns. The relations are defined using indexes and tables to speed up selected queries. And to query multiple tables simultaneously, it uses JOINS. NoSQL databases have different techniques to store data. Document databases may include nested documents and arrays while Graph databases either store edges and vertices or triples or quads etc. based on the type of data shape, we may choose the database.

10. Scope for multiple databases

There may be situations based on the data modeling w.r.t. to various sets of data structures, that more than one type of database may be required for the application. If choosing more than one databases, select one database that will own any specific set of data and this database acts as the canonical database for those entities. Additional databases that may work with this same set of data may have a copy of the data, but they will not be considered as the owner of this data.

11. Safety and security of data

What level of security of the data stored is being provided by a database is another important factor. If the data to be stored is highly confidential, a highly secured database is essential.

For security point of view, Authentication, Data in motion, Data at rest, Read/write access are some parameters to be considered. Also the safety measures implemented by the database, in case of any system crash or failure, is also important factor to consider while choosing a database.

12. Stability of database schemas

If the most fields to be consistent from record to record and database schemas remain unchanged, RDBMS or SQL databases are good. And if there may be variation in record to record data and also don't want any restrictions on the database schemas then, schema-free NoSQL databases will be a better option for the application.

13. Geographic distribution of users

With the globalization of activities and the geographic distribution of users is a very very important factor while choosing the right database. If database users are spread globally, the database latency will be lower unless additional servers are provided in their regions. Many databases supports distributed servers for both **read-write** while some offers **read-only** operations in distributed servers. And in case of read-only operations in distributed servers all the write operations are routed through the master server. The decision is touch and a trade-off between latency and consistency becomes harder due to geographic distribution.

14. Infrastructure requirements

There may be special infrastructure requirement for the database engine to run efficiently. It may be both hardware and software requirements.. Or it could be part of a larger data platform offering by a major cloud provider. Not only this, data migration is another point to consider here that what are the efforts required to move your data as needed.

15. Vendors

If we already have some vendors for infrastructure etc then the cost as well as time may be greatly reduced by leveraging an existing partnership. In other case, time, money, and efforts are required to bring on new vendor with the necessary skills for the chosen database engine.

16. Monitoring and Alerting

There is always a need to properly monitor the activities, raise alerts, and act promptly upon the alerts as necessary. This factor also need to be kept in mind that we want a complete

database solution or will deploy own monitoring solution. The factor is not only to consider the expenses, but the potential for needing to bring in expertise also.

17. ACID vs. BASE

RDBMS engines implements ACID (Atomicity, Consistency, Isolation, and Durability) properties very efficiently and effectively while NoSQL database engines uses BASE (Basic Availability, Soft-state, and Eventual consistency) to achieve the performance offered through horizontal scaling, high availability (HA), and fault tolerance. In very simple terminology relational systems tend to favor writes, whereas NoSQL systems tend to favor reads.

18. Programming language

The Programming language of the applications may also be a considerable factor that influences the choice of database. For example, for a JavaScript application, the database must support the JSON data type being the natural data format for JavaScript.

19. Budget and legality

Financial budget is a key factor in database selection because some databases are free, and some are paid versions. Not only this, we may also have to pay recurring support & maintenance charges in some cases. Moreover, how capable the database is in complying with the legal requirements is also a point to consider.

Assignment

1. What are various factors to consider while selecting the correct database?
2. Explain ACID and BASE w.r.t. SQL and NoSQL databases.