SQL and NOSQL: Which is Better

Mrs.Rohini G. Gaikwad, Assistant Professor, MCA Dept   **Dr.A.C.Goje, Director**
Research Scholar   **Research Guide**
Vidya Pratishthan's Institute Of Information Technology,   **Vidya Pratishthan's Institute Of Information Technology,**
Baramati,Pune,Maharashtra.   **Baramati,Pune,Maharashtra**

Abstract:

Most of you are already familiar with SQL database, and have a good knowledge on either MySQL, Oracle, or other SQL databases. In the last several years, NoSQL database is getting widely adopted to solve various business problems.

This paper presents differences between an SQL database management system - Oracle Database and NoSQL document oriented database management system - Mongo DB. The comparison criteria include theoretical differences, features, restrictions, distribution, Query evaluation.

Keyword: RDBMS,NOSQL,Operations

1] Introduction:

1.1 What Is NOSQL

NoSQL systems are distributed, non-relational data-bases designed for large-scale data storage and for massively-parallel data processing across a large number of commodity servers. They also use non-SQL languages and mechanisms to interact with data (though some new feature APIs that convert SQL queries to the system's na-tive query language or tool). NoSQL database systems arose alongside major Internet companies, such as Google, Amazon, and Facebook; which had challenges in dealing with huge quantities of data with conventional RDBMS solutions could not cope [1]. They can support multiple activities, including exploratory and predictive analytics, ETL-style data transformation, and non mis-sion-critical OLTP (for example, managing long-duration or inter-organization transactions). Originally motivated by Web 2.0 applications, these systems are designed to scale to thousands or millions of users doing updates as well as reads, in contrast to traditional DBMSs and data warehouses [2].

Most NoSQL databases are developed to run on clusters consisting of commodity computers and therefore have to be distributed and failure tolerant. To achieve this, they have to make different trade-offs regarding the ACID properties, transaction management, query capabilities and performance. They are usually designed to fit the requirements of most web services and most of them are schema free and bring their own query languages. *[3]*

1.2 Types Of NOSQL

The primary way in which NoSQL databases differ from relational databases is the data model. Although there are arguably dozens of NoSQL databases, they primarily fall into one of the following three categories.[4]

1.Document Model

Whereas relational databases store data in rows and columns, document databases store data in documents. These documents typically use a structure that is like JSON (JavaScript Object Notation), a format popular among developers. Documents provide an intuitive and natural way to model data that is closely aligned with object-oriented programming — each document is effectively an object. Documents contain one or more fields, where each field contains a typed value, such as a string, date, binary or array. Rather than spreading out a record across multiple columns and tables, each record and its associated data are typically stored together in a single document. This simplifies data access and reduces or even eliminates the need for joins and complex transactions.

In a document database, the notion of a schema is dynamic: each document can contain different fields. This flexibility can be particularly helpful for modeling unstructured and polymorphic data. It also makes it easier to evolve an application during development, such as adding new fields. Additionally, document databases generally provide the query robustness that developers have come to expect from relational databases. In particular, data can be queried based on any fields in a document.

Applications: Document databases are general purpose, useful for a wide variety of applications due to the flexibility of the data model, the ability to query on any field and the natural mapping of the document data model to objects in modern programming languages.

Examples: MongoDB and CouchDB.

2.Graph Model

Graph databases use graph structures with nodes, edges and properties to represent data. In essence, data is modeled as a network of relationships between specific elements. While the graph model may be counter-intuitive and takes some time to understand, it can be used broadly for a number of applications. Its main appeal is that it makes it easier to model relationships between entities in an application.

Applications: Graph databases are useful in cases where relationships are core to the application, like social networks.

Examples: Neo4j and HyperGraphDB.

3.Key-Value and Wide Column Models

From a data model perspective, key-value stores are the most basic type of NoSQL database. Every item in the database is stored as an attribute name, or key, together with its value. The value, however, is entirely opaque to the system; data can only be queried by the key. This model can be useful for representing polymorphic and unstructured data, as the database does not enforce a set schema across key-value pairs.

Wide column stores, or column family stores, use a sparse, distributed multi-dimensional sorted map to store data.

Each record can vary in the number of columns that are stored, and columns can be nested inside other columns called super columns. Columns can be grouped together for access in column families, or columns can be spread across multiple column families. Data is retrieved by primary key per column family.

Applications: Key value stores and wide column stores are useful for a narrow set of applications that only query data by a single key value. The appeal of these systems is their performance and scalability, which can be

highly optimized due to the simplicity of the data access patterns. Examples: Riak and Redis (Key-Value); HBase and Cassandra (Wide Column).

1.3] Oracle-

 Oracle Database (commonly referred to as Oracle RDBMS or simply as Oracle) is an object-relational database management system produced and marketed by Oracle Corporation.

As a database system, RDBMS stores data in a database object called table. Table is a structure of dataset consist of row and column. In order to access the data stored in RDBMS, SQL language is used. The SQL use statement: SELECT, INSERT, UPDATE, DELETE to manipulate the data stored in database system.[5]

RDBMS and its SQL language has been an industrial standard database system, since 1986 ANSI (American National Standard of Industry) issued an SQL standard. Then in 1987, International Standard of Organization (ISO) issued SQL standard as International standard as ISO/IEC 9075. However, in many RDBMS, there are variations of SQL from the ISO and ANSI standard, which unique to each database. However, the basic standard is the same in every database using SQL.[5]

 1.4] MongoDB-

The idea of NoSQL started in 1998. The name is a database created by Carlo Strozzi, who made its database management system and use a stream-operator paradigm to access data instead of SQL syntax, Stream operator paradigm is a database operator which resemble a mathematical operation in order to access data stored in database.

MongoDB is an open-source document database that provides high performance, high availability, and automatic scaling.

MongoDB was developed by 10gen and was initially released in 2009. It was developed using C++. It is a high performance and efficient database. It provides features like consistency fault tolerance, persistence. MongoDB provides additional features like aggregation, ad hoc queries, indexing, auto sharding etc. In MongoDB the documents are mainly stored in BSON (Binary JSON) format. BSON documents contain an ordered list of elements consisting of field name, type and value. BSON is efficient both in storage space and scan speed when compared to JSON. MongoDB uses GridFS as a specification for storing large files. MongoDB is well suited for applications like content management systems, archiving, real time analytics etc. MongoDB is currently being used by MTV networks, Foursquare, The Guardian etc. It is also being used in projects like CERN"s LHC, UIDAI Aadhaar which is India's unique identification project. The disadvantages are that it can be unreliable and indexing takes up lot of RAM[6][7].

MongoDB is the world's most popular NoSQL database,powering tens of thousands of operational and analytical big data applications in mission-critical financial services,telecoms, government and enterprise environments .[12]

2] Basic Terminology:

No	Oracle	MongoDB
1.	Table	Collection
2.	Record/Tuple	Document
3.	Primary Key Constraint	Object ID

Table 1:Basic terminology

3] Experimental Setup:

For this experiment machine installed with Oracle version : Oracle 11g and MongoDB version : Mongodb 2.6.7

4] SQL Vs NOSQL

The NoSQL systems described here generally do not provide ACID transactional properties: updates are eventually propagated, but there are limited guarantees
on the consistency of reads. Some authors suggest a "BASE" acronym in contrast to the "ACID" acronym:
• BASE = Basically Available, Soft state,
Eventually consistent
• ACID = Atomicity, Consistency, Isolation, and Durability

The idea is that by giving up ACID constraints, one can achieve much higher performance and scalability.
However, the systems differ in how much they give up. For example, most of the systems call themselves "eventually consistent", meaning that updates are eventually propagated to all nodes, but many of them provide mechanisms for some degree of consistency, such as multi-version concurrency control (MVCC).[8]
SQL (relational) versus NoSQL scalability is a divisive topic. Here is some more background to support this position.

Operation	Oracle	MongoDB
Create	Create Database	Use DBNAME
Create	Create Table	Db.createCollection
Add	Insert Into	Db.Collection.insert
Update	Update table	Db.collection.update
Retrive	Select	Db.collection.find()

Table 2: Commands for operations

For conducting the experiment in Oracle Employee database is created .In Mongodb employee collection is created.

a) Operators

Oracle: supports logical,conditional,relational operators.

Select * from employee where salary>= 10000 and <20000.

MongoDB: Rather than symbolic operators supports operator name with $

For example $gte: greater than

db.employee.find({salary:{$gte :10000,$lte :20000}})

b) Importing a file in to Database

 Oracle:

In oracle we need use Import utility in SQL Loader .First create a table emp_info having same fields as csv file.
Then write a code in notepad

```
 LOAD DATA
INFILE 'employee.csv'
INTO TABLE emp_info
FIELDS TERMINATED BY "," ;
```

Run batch file in command prompt.

MongoDB : MongoDB provides mongoimport utility for importing .csv and Json file.

 Mongoimport –d mydata   -c employee --type csv -- file emp.csv --headerline

c) Execution Statistics

*Oracle:*In Oracle DBMS_STATS has procedures as

Statistics Gathering Procedures in the DBMS_STATS Package

Procedure	Collects
GATHER_INDEX_STATS	Index statistics
GATHER_TABLE_STATS	Table, column, and index statistics
GATHER_SCHEMA_STATS	Statistics for all objects in a schema
GATHER_DICTIONARY_STATS	Statistics for all dictionary objects
GATHER_DATABASE_STATS	Statistics for all objects in a database

MongoDB :

*d*b.employee.find({}).explain()

db.employee.find({}).explain(executionStats)

d) Sharding:

Sharding, or *horizontal scaling*, by contrast, divides the data set and distributes the data over multiple servers, or **shards**. Each shard is an independent database, and collectively, the shards make up a single logical database

In oracle scaling is vertical. In essence, more data means a bigger server, which can get very expensive. It is possible to scale an RDBMS across multiple servers, but this is a difficult and time-consuming process.
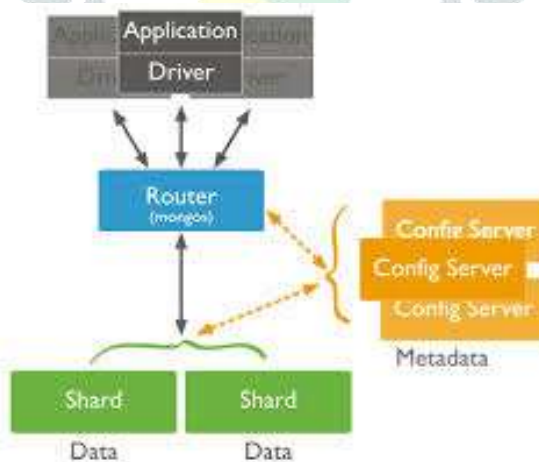
Fig1: Sharding in Oracle[11]
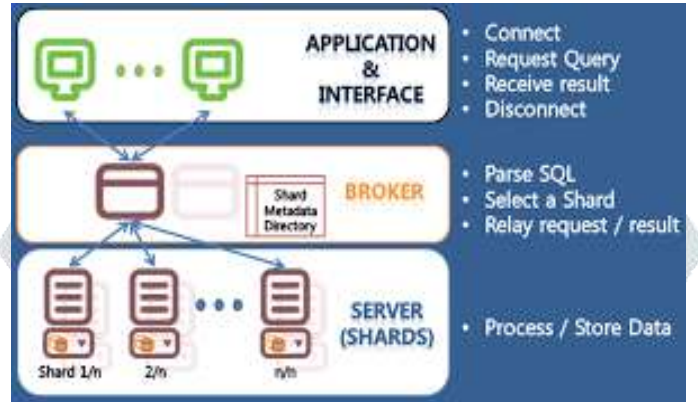
MongoDB supports sharding through the configuration of a *sharded clusters*.

Sharding reduces the number of operations each shard handles. Sharding reduces the amount of data that each server needs to store. Each shard stores less data as the cluster grows.

Fig2: Sharding in MongoDB [7]

e) MapReduce

Map-reduce is a data processing paradigm for condensing large volumes of data into useful *aggregated* results. For map-reduce operations, MongoDB provides the mapReduce database command.

Oracle:

Oracle Table Functions are a proven technology, used by many internal and external parties to extend Oracle Database 11g.

Oracle Table Functions are a robust scalable way to implement Map-Reduce within the Oracle database and leverage the scalability of the Oracle Parallel Execution framework. Using this in combination with SQL provides an efficient and simple mechanism for database developers to develop Map-Reduce functionality within the environment they understand and with the languages they know.[9]

MongoDB:

All map-reduce functions in MongoDB are JavaScript and run within the mongod process. Map-reduce operations take the documents of a single *collection* as the *input* and can perform any arbitrary sorting and limiting before beginning the map stage. mapReduce can return the results of a map-reduce operation as a document, or may write the results to collections. The input and the output collections may be sharded.

Conclusion:

The idea that SQL and NoSQL are in direct opposition and competition with each other is flawed one, not in the least because many companies opt to use them concurrently.

In this Paper ,Researcher outline the differences between SQL and NoSQL, the vast array of differences within NoSQL technologies themselves as:

- Data Storage:

  Oracle Stores the data in terms of relational model which has rows and columns. It has fixed schema structure.

  MongoDB stores data as document in colletions in JSON format.It has schemaless structure.

- Query Operations: Oracle supports DDL,DML operations for data manipulation. MongoDB supports Curd operations for data manipulation less restrictions.
- Statistics
  Oracle has packages like DBMS_STATS include procedures for generating execution statistics whereas MongoDB has functions like explain().

  - Scaling:
    In Oracle scaling, you would start with large server and scale up until you have maximized the   computing resources.
    In mongoDB scaling is horizontal, meaning   across servers.

- While Importing in Oracle we need to follow certain procedure . In Mongodb single line command is sufficient.

- The majority of relational databases are ACID compliant. But many NoSQL solutions sacrifice ACID compliancy for performance and scalability.

Although NOSQL has evolved at a very high rapidity, it still lags behind relational database since it lacks a standard query language. If a standard query language for NOSQL is introduced, it will surely benefit to majority users. Also there is no fixed course of action for choosing the NOSQL database for enterprise. Users must first consider various parameters like query language, the interface, availability, redundancy, consistency and analyze the pros and cons of various data models before choosing a particular data model.

Final words, if you want to use a fast, flexible database, you can rely on NoSQL Databse. If the alacrity of the database isn't main concern, and if you need relations between the tables and the collections, you can go with traditional of RDBMS.

Acknowledgement:

References:

[1] http://en.wikipedia.org/wiki/NoSQL

[2] R. Cattell, (2010) "Scalable SQL and NoSQL Data Stores," ACM SIG-MODRecord, vol. 39.

[3] *"Analysis and Classi_cation of NoSQL Databases   and Evaluation of their Ability to Replace an Object-relationalPersistence Layer" Author: Kai Orend Supervisor: Prof. Florian Matthes, Ph.D.Advisor: Thomas Büchner, Ph.D   Submission Date: 14.04.2010*

[4]"Top 5 Considerations When Evaluating NoSQL Databases February 2015"

[5] "MongoDB vs MySQL: Comparison Between RDBMS and Document Oriented Database "   MAY 2, 2014 BY KORNELIUSZ

[6] "Type of NOSQL Databases and its Comparison with Relational Databases " Ameya Nayak ,Anil Poriya , Dikshay Poojary Dept. of Computer Engineering Thakur College of Engineering and Technology .University of Mumbai . *International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 5– No.4, March 2013 – www.ijais.org*

 [7]www.mongodb.org

[8] *Scalable SQL and NoSQL Data Stores Rick Cattell Originally published in 2010, last revised December 2011*

[9] Oracle White Paper— In-database Map-Reduce 2009

 [10] http://dataconomy.com/sql-vs-nosql-need-know/ MongoDB vs Oracle -- Database Comparison Alexandru Boicea ,Florin Radulescu ,Laura Ioana Agapin   DOI Bookmark: http://doi.ieeecomputersociety.org/10.1109/EIDWT.2012.32

[11]http://www.dba-oracle.com/t_sharding.html

[12] MongoDB Operations Best Practices MongoDB 2.6 October 2014.

 [13]MongoDB vs Oracle - database comparison Alexandru Boicea, Florin Radulescu, Laura Ioana Agapin Faculty of Automatic Control and Computer Science Politehnica University of Bucharest Bucharest, Romania alexandru.boicea@cs.pub.ro, florin.radulescu@cs.pub.ro, lauraioana.agapin@gmail.com