

Assignment No.: 1

Date:

Title: -

Study of Open source NOSQL Databases and Compare the different database systems based on points like efficiency, scalability, characteristics and performance.
uitable exale

Aim: Study of Open source NOSQL Databases and Compare the different database systems based on points like efficiency, scalability, characteristics and performance.

Objective: To be aware different Open Source Databases.

Theory:

1) A brief description of Open Source Database:

Open Source Software:

Open-source software (OSS) is computer software with its source code made available and licensed with a license in which the copyright holder provides the rights to study change and distribute the software to anyone and for any purpose.

Open-source software is very often developed in a public, collaborative manner. Open-source software is the most prominent example of open-source development and often compared to (technically defined) user-generated content or (legally defined) open-content movements.

A database is a base for data. An Open Source database is a base for data that includes Free and Open Source Software. Open source software is software that makes the source code available to anyone. The user is allowed to implement, share and further develop the database software to suit various needs.

Popular Open Source Databases

Following are open source databases

. 1) MySQL

MySQL "My S-Q-L", officially, but also called "My Sequel" is the world's most widely used open-source relational database management system (RDBMS). It is named after co-founder Michael Widenius's daughter, My. The SQL phrase stands for Structured Query Language. MySQL is a popular choice of database for use in web applications, and is a central

component of the widely used LAMP open source web application software stack. LAMP is an acronym for "Linux, Apache, MySQL, Perl/PHP/Python." Free-software-open source projects that require a full-featured database management system often use MySQL.

MySQL is also used in many high-profile, large-scale websites, including Wikipedia, Google (though not for searches), Facebook, Twitter, Flickr, and YouTube. MySQL is the most popular and widely used relational database management system that provides multi-user access to number of databases. MySQL is now owned by Oracle and uses Sequential Query Language to manage database. Its source is available under GNU license and propriety agreements. MySQL is most popular among PHP developers and used for websites, web applications and online services.

Features of MySQL:

- Because of its unique storage engine architecture MySQL performance is very high.
- Supports large number of embedded applications which makes MySQL very flexible.
- Use of Triggers, Stored procedures and views which allows the developer to give a higher productivity.
- Allows transactions to be rolled back, commit and crash recovery
- Embedded database library
- Full-text indexing and searching
- Updatable views
- Cursors
- Triggers
- Cross-platform support

Limitation of MySQL:

- Like other SQL databases, MySQL does not currently comply with the full SQL standard for some of the implemented functionality, including foreign key references when using some storage engines other than the default of InnoDB
- No triggers can be defined on views.

- MySQL, like most other transactional relational databases, is strongly limited by hard disk performance. This is especially true in terms of write latency.

2) PostgreSQL

It is developed by PostgreSQL Global Development Group and is an ORDBMS (Object Relational Database Management System). Available for all platforms Mac, Windows, Solaris and Linux under MIT license, PostgreSQL supports all the properties of major databases. PostgreSQL is currently available as version 9.1.

3) SQLite

SQLite is a small lightweight embedded database used in Application File formats, Database for mobile apps and websites. SQLite has compliance with ACID properties of database. It is faster and has simple to use API. SQLite comes with a standalone command-line interface (CLI) client that can be used to administer SQLite databases.

4) Berkeley DB

Owned by Oracle, Berkeley DB provides the foundational storage services for your application, no matter how demanding and unique your requirements may seem to be. Berkeley DB APIs are available in almost all programming languages including ANSI-C, C++, Java, C#, Perl, Python, Ruby and Erlang.

A program accessing the database is free to decide how the data is to be stored in a record. Berkeley DB puts no constraints on the record's data. The record and its key can both be up to four gigabytes long. Note that Berkeley DB is not a full DBMS.

5) Firebird

Firebird has always been more fully featured than MySQL, and has, unlike PostgreSQL, always worked well on Windows as well as Linux and other 'Nix variants. Firebird provides a lot of the

features available in commercial databases, including stored procedures, triggers, hot backups (backups while the database is running) and replication. Firebird database comes in two variations, classic server and super server.

6) MongoDB

MongoDB is a cross-platform document-oriented database. Classified as a NoSQL database, MongoDB eschews the traditional table-based relational database structure in favor of JSON-like documents with dynamic schemas (MongoDB calls the format BSON), making the integration of data in certain types of applications easier and faster. Released under a combination of the GNU Affero General Public License and the Apache License, MongoDB is free and open-source software.

Development of MongoDB began in 2007, when the company (then named 10gen) was building a platform as a service similar to Windows Azure or Google App Engine. In 2009, MongoDB was open sourced as a stand-alone product with an AGPL license. MongoDB has been adopted as backend software by a number of major websites and services, including eBay, Foursquare, SourceForge, Viacom, and the New York Times, among others. MongoDB is the most popular NoSQL database system.

Some of the main features include:

Ad hoc queries

MongoDB supports search by field, range queries, regular expression searches. Queries can return specific fields of documents and also include user-defined JavaScript functions.

Indexing

Any field in a MongoDB document can be indexed (indices in MongoDB are conceptually similar to those in RDBMSes). Secondary indices are also available.

Replication

MongoDB provides high availability with replica sets. A replica set consists of two or more copies of the data. Each replica set member may act in the role of primary or secondary replica at any time. The primary replica performs all writes and reads by default. Secondary replicas maintain a copy of the data on the primary using built-in replication. When a primary replica fails, the replica set automatically conducts an election process to determine which secondary should become the primary. Secondaries can also perform read operations, but the data is eventually consistent by default.

Load balancing

MongoDB scales horizontally using sharding. The user chooses a shard key, which determines how the data in a collection will be distributed. The data is split into ranges (based on the shard key) and distributed across multiple shards. (A shard is a master with one or more slaves.) MongoDB can run over multiple servers, balancing the load and/or duplicating data to keep the system up and running in case of hardware failure. Automatic configuration is easy to deploy, and new machines can be added to a running database.

File storage

MongoDB can be used as a file system, taking advantage of load balancing and data replication features over multiple machines for storing files.

This function, called GridFS, is included with MongoDB drivers and available with no difficulty for development languages. MongoDB exposes functions for file manipulation and content to developers. GridFS is used, for example, in plugins for NGINX and lighttpd. Instead of storing a file in a single document, GridFS divides a file into parts, or chunks, and stores each of those chunks as a separate document. In a multi-machine MongoDB system, files can be distributed and copied multiple times between machines transparently, thus effectively creating a load-balanced and fault-tolerant system.

Aggregation

Map Reduce can be used for batch processing of data and aggregation operations. The aggregation framework enables users to obtain the kind of results for which the SQL GROUP BY clause is used.

Server-side JavaScript execution

JavaScript can be used in queries, aggregation functions (such as MapReduce), and sent directly to the database to be executed.

Capped collections

MongoDB supports fixed-size collections called capped collections. This type of collection maintains insertion order and, once the specified size has been reached, behaves like a circular queue.

Limitations of MongoDB:

- On 32-bit, it has limitation of 2.5 Gb data
- 4 MB/16 MB document size limitation depending on version
- Read/write lock is currently global level
- No joins across collections
- No transaction support
- No referential integrity support
- Need to have enough memory to fit your working set into memory.

7) CouchDB

Apache CouchDB, commonly referred to as **CouchDB**, is an open source database that focuses on ease of use and on being "a database that completely embraces the web". It is a NoSQL database that uses JSON to store data, JavaScript as its query language using Map

Reduce, and HTTP for an API. One of its distinguishing features is multi-master replication. CouchDB was first released in 2005 and later became an Apache project in 2008.

Unlike in a relational database, CouchDB does not store data and relationships in tables. Instead, each database is a collection of independent documents. Each document maintains its own data and self-contained schema. An application may access multiple databases, such as one stored on a user's mobile phone and another on a server. Document metadata contains revision information, making it possible to merge any differences that may have occurred while the databases were disconnected.

CouchDB implements a form of Multi-Version Concurrency Control (MVCC) in order to avoid the need to lock the database file during writes. Conflicts are left to the application to resolve. Resolving a conflict generally involves first merging data into one of the documents, then deleting the stale one.

CouchDB (*Couch* is an acronym for *cluster of unreliable commodity hardware*) is a project created in April 2005 by Damien Katz, former Lotus Notes developer at IBM. Damien Katz defined it as a "storage system for a large scale object database". His objectives for the database were to become the database of the Internet and that it would be designed from the ground up to serve web applications. He self-funded the project for almost two years and released it as an open source project under the GNU General Public License.

In February 2008, it became an Apache Incubator project and the license was changed to the Apache License. A few months after, it graduated to a top-level project. This led to the first stable version being released in July 2010. In early 2012, Damien Katz left the project to focus on Server. Since the departure of Damien Katz, the Apache CouchDB project has continued, releasing 1.2 in April 2012 and 1.3 in April 2013. In July 2013, the CouchDB community merged the codebase for Big Couch, Cloudant's clustered version of CouchDB, into the Apache project. The BigCouch clustering framework is prepared to be included in an upcoming release of Apache CouchDB.

System Properties Comparison CouchDB vs. MongoDB vs. MySQL

Name	CouchDB	MongoDB	MySQL
Description	A document store inspired by Lotus Notes	One of the most popular document stores	Widely used open source RDBMS
Developer	Apache Software Foundation	MongoDB, Inc	Oracle
Initial release	2005	2009	1995
License	Open Source	Open Source	Open Source
Implementation language	Erlang	C++	C and C++
Database model	Document store	Document store	Relational DBMS
Data scheme	schema-free	schema-free	yes
SQL	no	no	yes
Supported programming languages	C C# Java JavaScript Lisp	C C# Java JavaScript Lisp	C C# Java PHP Python
Triggers	yes	no	Yes
Foreign keys	no	no	yes
Map Reduce	yes	yes	no

Conclusion: Awareness of Different Open Source Databases.

FAQs: -

1. What is Open Source Software?
2. What is Open Source Database?
3. What are different open source databases?
4. What is difference between MySQL and MongoDB?
5. Foreign key is possible in MongoDB or Not?
6. List more open-source database systems other than MySQL and MongoDB?

Assignment No. 2

Date:

Title: Install and configure client and server for MySQL and MongoDB (Show all commands and necessary steps for installation and configuration).

Aim: Install and configure client and server for MySQL and MongoDB

Objective: Study all commands and necessary steps for installation and configuration.

Theory:

Setting Up the MySQL Database Server in the Windows Operating System

- Starting the Download
- Starting the Installation

Starting the Download

1. Go to <http://dev.mysql.com/downloads/installer/>.
2. Click the Download button.
3. Save the installer file to your system.

Starting the Installation

After the download completes, run the installer as follows:

1. Right-click the downloaded installation file (for example, mysql-installer-community-5.6.14.0.msi) and click Run.
The MySQL Installer starts.
2. On the Welcome panel, select Install MySQL Products.
3. On the License Information panel, review the license agreement, click the acceptance checkbox, and click Next.
4. On the Find latest products panel, click Execute.
When the operation is complete, click Next.
5. On the Setup Type panel, choose the Custom option and click Next.
6. On the Feature Selection panel, ensure MySQL Server 5.6.x is selected, and click Next.
7. On the Check Requirements panel, click Next.
8. On the Installation panel, click Execute.
When the server installation is completed successfully, the information message appears on the Installation panel. Click Next.
9. On the Configuration panel, click Next.

10. At the first MySQL Server Configuration page (1/3), set the following options:

- **Server Configuration Type.** Select the Development Machine option.
- **Enable TCP/IP Networking.** Ensure the checkbox is selected and specify the options below:
- **Port Number.** Specify the connection port. The default setting is 3306 - leave it unchanged if there is not special reason to change it.
- **Open Firewall port for network access.** Select to add firewall exception for the specified port.
- **Advanced Configuration.** Select the Show Advanced Options checkbox to display an additional configuration page for setting advanced options for the server instance if required.

Note: Choosing this option is necessary to get to the panel for setting the network options where you will turn off the firewall for the port used by the MySQL server.

11. Click Next.

12. At the second MySQL Server Configuration page (2/3), set the following options:

- **Root Account Password.**
- **MySQL Root Password.** Enter the root user's password.
- **Repeat Password.** Retype the root user's password.

Note: The root user is a user who has full access to the MySQL database server - creating, updating, and removing users, and so on. Remember the root password - you will need it later when creating a sample database.

- **MySQL User Accounts.** Click Add User to create a user account. In the MySQL User Details dialog box, enter a user name, a database role, and a password (for example, !phpuser). Click OK.

Click Next.

13. At the third MySQL Server Configuration page (3/3), set the following options:

- **Windows Service Name.** Specify a Windows Service Name to be used for the MySQL server instance.
- **Start the MySQL Server at System Startup.** Leave the checkbox selected if the MySQL server is required to automatically start at system startup time.
- **Run Windows Service as.** Choose either:
 - **Standard System Account.** Recommended for most scenarios.
 - **Custom User.** An existing user account recommended for advanced scenarios.

Click Next.

14. At the Configuration Overview page, click Next.

15. When the configuration is completed successfully, the information message appears on the Complete panel. Click Finish.

Note: To check that the installation has completed successfully, run the Task Manager. If the MySQLd-nt.exe is on the Processes list - the database server is running.

Installation Steps for MongoDB.

At Server side:

1) Extract Zip File

2) C:\Users\admin>cd E:\mongodb-win32-x86_64-2008plus-2.6.2\mongodb-win32-x86_64-2008plus-2.6.2\bin

Ex: E:\Teacher (this is the folder which contains information related to Teacher. Here we are creating Teacher Database. Which contain the information of Teacher_id, name of a teacher, department of a teacher, salary and status of a teacher. Here status is wheather teacher is approved by the university or not. Our main idea is to implement all the DDL & DML queries on the Teacher Database.

3. C:\Users\admin>E:

4. E:\mongodb-win32-x86_64-2008plus-2.6.2\mongodb-win32-x86_64-2008plus-2.6.2\bin>mongod.exe --dbpath E:\Teacher

Note: keep the server in running state.

At Client Side:

- Open Another CMD prompt
- Go to bin folder of MongoDB

1. C:\Users\admin>cd E:\mongodb-win32-x86_64-2008plus-2.6.2\mongodb-win32-x86_64-2008plus-2.6.2\bin

2. C:\Users\admin>E:

3. E:\mongodb-win32-x86_64-2008plus-2.6.2\mongodb-win32-x86_64-2008plus-2.6.2\bin>mongo.exe Teacher

4. MongoDB shell version: 2.6.2

connecting to: Teacher

Now Teacher database is ready. U can perform all the related operations on the teacher database.

At server side:

U will find the following:

2014-06-20T17:44:09.233+0530 [initandlisten] connection accepted from 127.0.0.1:49360 #1 (1 connection now open)

Now it indicates both server and client are ready.