

MONGODB - THE NOSQL DATABASE

Akhil Latta

Software Engineer

Z Systems, Mohali, Punjab

MongoDB is an open source document-oriented database system developed and supported by 10gen. It is part of the NoSQL family of database systems. Instead of storing data in tables as is done in a "classical" relational database, MongoDB stores structured data as JSON-like documents with dynamic schemas (MongoDB calls the format BSON), making the integration of data in certain types of applications easier and faster. 10gen began development of MongoDB in October 2007. The database is used by eBay, MetLife, Telefónica, Foursquare, MTV Networks and the UK Government. MongoDB is the most popular NoSQL database management system. Binaries are available for Windows, Linux, OS X, and Solaris. This paper highlights the emerging database engine mongodb and its applications.

INTRODUCTION

Development of MongoDB began at 10gen in 2007, when the company 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.

In March 2010, from version 1.4, MongoDB has been considered production ready. The latest stable version, 2.4.5, was released in July 3 2013.

LICENSING AND SUPPORT

MongoDB is available for free under the GNU Affero General Public License. The language drivers are available under an Apache License. In addition, 10gen offers commercial licenses for MongoDB.

MAIN FEATURES

The following is a brief summary of some of the main features:

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 supports master-slave replication. A master can perform reads and writes. A slave copies data from the master and can only be used for reads or backup (not writes). The slaves have the ability to select a new master if the current one goes down.

Load balancing

MongoDB scales horizontally using sharding. The developer 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 could 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

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

MapReduce 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), are 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.

LANGUAGE SUPPORT

MongoDB has official drivers for a variety of popular programming languages and development environments. Web programming language Opa also has built-in support for MongoDB, which is tightly integrated in the language and offers a type-safety layer on top of

MongoDB. There are also a large number of unofficial or community-supported drivers for other programming languages and frameworks.

HTTP/REST INTERFACES

There are REST and HTTP interfaces that allow the manipulation of MongoDB entries via HTTP GET, POST, UPDATE, and DELETE calls.

MANAGEMENT AND GRAPHICAL FRONT-ENDS

MONGODB TOOLS

In a MongoDB installation the following commands are available:

mongo

MongoDB offers an interactive shell called **mongo**, which lets developers view, insert, remove, and update data in their databases, as well as get replication information, set up sharding, shut down servers, execute JavaScript, and more.

Administrative information can also be accessed through a **web interface**, a simple webpage that serves information about the current server status. By default, this interface is 1000 ports above the database port (28017).

mongostat

mongostat is a command-line tool that displays a summary list of status statistics for a currently running MongoDB instance: how many inserts, updates, removes, queries, and commands were performed, as well as what percentage of the time the database was locked and how much memory it is using. This tool is similar to the UNIX/Linux vmstat utility.

mongotop

mongotop is a command-line tool providing a method to track the amount of time a MongoDB instance spends reading and writing data. mongotop provides statistics on

the per-collection level. By default, mongotop returns values every second. This tool is similar to the UNIX/Linux top utility.

mongosniff

mongosniff is a command-line tool providing a low-level tracing/sniffing view into database activity by monitoring (or "sniffing") network traffic going to and from MongoDB. mongosniff requires the Libpcap network library and is only available for Unix-like systems. A cross-platform alternative is the open source Wireshark packet analyzer which has full support for the MongoDB wire protocol.

mongoimport, mongoexport

mongoimport is a command-line utility to import content from a JSON, CSV, or TSV export created by mongoexport or potentially other third-party data exports. Usage information can be found in the MongoDB Manual's section on Importing and Exporting MongoDB Data.

mongodump, mongorestore

mongodump is a command-line utility for creating a binary export of the contents of a Mongo database; mongorestore can be used to reload a database dump. Data backup strategies and considerations are detailed in the MongoDB Manual's section on Backup and Restoration Strategies.

MONITORING PLUGINS

There are MongoDB monitoring plugins available for the following network tools:

- mongo-munin: Plugin for Munin
- mongodb-ganglia: Plugin for ganglia
- MongoDB Cacti Graphs: Plugin for cacti
- MongoDB Slow Queries: Plugin for Scout

More monitoring and diagnostic tools for MongoDB are listed on MongoDB Admin Zone: [Monitoring and Diagnostics](#)

CLOUD-BASED MONITORING SERVICES

- MongoDB Management Service (MMS) is a cloud-based suite of services for managing MongoDB deployments. It includes monitoring and backup capabilities. MMS is developed by 10gen, the company behind MongoDB.
- Server Density is a cloud-based tool which helps you provision and monitor your infrastructure. It includes a custom dashboard for MongoDB, MongoDB specific alerts, replication failover timeline and iPhone, iPad and Android mobile apps.

WEB AND DESKTOP APPLICATION GUIS

Several GUIs have been created by MongoDB's developer community to help visualize their data. Some popular ones are:

OPEN SOURCE TOOLS

- RockMongo: PHP-based MongoDB administration GUI tool
- phpMoAdmin: another PHP GUI that runs entirely from a single 95kb self-configuring file
- UMongo: a desktop application for all platforms.
- Mongo3: a Ruby-based interface.
- Meclipse: Eclipse plugin for interacting with MongoDB
- MongoHub: a freeware native Mac OS X application for managing MongoDB. Version for other operating systems is built on Titanium Desktop.
- mViewer: A simple web-based Administration and Management Tool for MongoDB written in Java.
- MongoDBPumper: a commercial high-performance data transfer solution to provide export and import functionality between Oracle and MongoDB databases.

More client tools for MongoDB are listed on MongoDB Administrator Manual

BUSINESS INTELLIGENCE TOOLS AND SOLUTIONS

- Jaspersoft BI: Java based Report Designer and Report Server that supports MongoDB
- Pentaho: MongoDB connectors for Pentaho Kettle and Pentaho BI
- RJMetrics: A hosted Business Intelligence Solution that supports MongoDB.
- eCommerce Analytics: eCommerce Analytics Software that supports MongoDB data analysis.
- Nucleon BI Studio: MS Windows based business intelligence software that supports MongoDB and other RDBMS.

REFERENCES

1. DB-Engines Ranking
2. MongoDB daddy: My baby beats Google BigTable
3. ^{a b} The MongoDB NoSQL Database Blog, The AGPL
4. The MongoDB NoSQL Database Blog, MongoDB 1.4 Ready for Production
5. MongoDB Support by 10gen
6. Article "Sharding" on MongoDB Administrator's Manual
7. GridFS article on MongoDB Developer's Manual
8. NGINX plugin for MongoDB source code
9. lighttpd plugin for MongoDB source code.
10. "MongoDB Drivers and Client Libraries — MongoDB Ecosystem 2.2.2".
Mongodb.org. Retrieved 2013-07-08.