

# Performance Best Practices for MongoDB

MongoDB 3.0  
February 2015

# Table of Contents

Introduction	1
Pluggable Storage Engines	1
Hardware	2
Application Patterns	3
Schema Design & Indexes	4
Disk I/O	5
Considerations for Amazon EC2	5
Considerations for Benchmarks	5

# Introduction

MongoDB is a high-performance, scalable database designed for a broad array of modern applications. It is used by organizations of all sizes to power online applications where low latency, high throughput and continuous availability are critical requirements of the system.

This guide outlines considerations for achieving performance at scale in a MongoDB system across a number of key areas, including hardware, application patterns, schema design and indexing, disk I/O, Amazon EC2, and designing for benchmarks. While this guide is broad in scope, it is not exhaustive. Following the recommendations in this guide will reduce the likelihood of encountering common performance limitations, but it does not guarantee good performance in your application.

MongoDB works closely with users to help them optimize their systems. Users should monitor their systems to identify bottlenecks and limitations. There are a variety of tools available, the most comprehensive of which is MongoDB Ops Manager, discussed later in this guide.

For a discussion on the architecture of MongoDB and some of its underlying assumptions, see the [MongoDB Architecture Guide](#). For a discussion on operating a MongoDB system, see the [MongoDB Operations Best Practices](#).

## MongoDB Pluggable Storage Engines

MongoDB 3.0 ships with two supported storage engines:

- The default MMAPv1 engine, an improved version of the engine used in prior MongoDB releases.
- The new WiredTiger storage engine. For many applications, WiredTiger's more granular concurrency control and native compression will provide significant benefits in the areas of lower storage costs, greater hardware utilization, and more predictable performance.

Both storage engines can coexist within a single MongoDB replica set, making it easy to evaluate and migrate between

them. Upgrades to the WiredTiger storage engine are non-disruptive for existing replica set deployments; applications will be 100% compatible, and migrations can be performed with zero downtime through a rolling upgrade of the MongoDB replica set. WiredTiger is enabled by starting the server using the following option:

```
mongod --storageEngine wiredTiger
```

Review the [documentation](#) for a checklist and full instructions on the migration process.

While each storage engine is optimized for different workloads users still leverage the same MongoDB query language, data model, scaling, security and operational tooling independent of the engine they use. As a result most of best practices in this guide apply to both supported storage engines. Any differences in recommendations between the two storage engines are noted.

## Hardware

MongoDB is designed for horizontal scale out. Rather than scaling up with ever larger and more expensive servers, users are instead encouraged to scale their systems by using many commodity servers operating together as a cluster. MongoDB provides native replication to ensure availability; auto-sharding to uniformly distribute data across servers; and in-memory computing to provide high performance without resorting to a separate caching layer. The following considerations will help you optimize the hardware of your MongoDB system.

**Ensure your working set fits in RAM.** As with most databases, MongoDB performs best when the working set (indexes and most frequently accessed data) fits in RAM. Sufficient RAM is the most important factor for hardware; other optimizations may not significantly improve the performance of the system if there is insufficient RAM. If your working set exceeds the RAM of a single server, consider sharding your system across multiple servers. Use the `serverStatus` command to view an estimate of the the current working set size.

**Use SSDs for write-heavy applications.** Because most disk I/O in MongoDB is random, SSDs can provide a significant performance improvement for write-heavy

systems. Data access is dominated by seek time in disk subsystems, and for spinning disks seek times tend to be around ~5 ms, while for SSD seek times tend to be around 0.1 ms, around 50x faster than spinning disks. SSD can also mitigate the latencies associated with working sets that do not fit in RAM, however in general it is best to ensure your working set fits in RAM.

**Configure compression for storage and I/O-intensive workloads.** MongoDB natively supports compression when using the WiredTiger storage engine. Compression reduces storage footprint by as much as 80%, and enables higher storage I/O scalability as fewer bits are read from disk. As with any compression algorithm administrators trade storage efficiency for CPU overhead, and so it is important to test the impacts of compression in your own environment.

MongoDB offers administrators a range of compression options for both documents and indexes. The default snappy compression algorithm provides a balance between high document and journal compression ratio (typically around 70%, dependent on data types) with low CPU overhead, while the optional zlib library will achieve higher compression, but incur additional CPU cycles as data is written to and read from disk. Indexes use prefix compression by default, which serves to reduce the in-memory footprint of index storage, freeing up more of the working set for frequently accessed documents. Testing has shown a typical 50% compression ratio using the prefix algorithm, though users are advised to test with their own data sets. Administrators can modify the default compression settings for all collections and indexes. Compression is also configurable on a per-collection and per-index basis during collection and index creation.

**Allocate CPU hardware budget for faster CPUs.**

MongoDB will deliver better performance on faster CPUs. The MongoDB WiredTiger storage engine is better able to saturate multi-core processor resources than the MMAPv1 storage engine.

**Dedicate each server to a single role in the system.**

For best performance, users should run one `mongod` process per host. With appropriate sizing and resource allocation using virtualization or container technologies, multiple MongoDB processes can run on a single server without contending for resource. If using the WiredTiger

storage engine, administrators will need to calculate the appropriate cache size for each instance by evaluating what portion of total RAM each of them should use, and splitting the default `cache_size` between each.

For resilience, multiple members of the same replica set should not be co-located on the same physical hardware.

**Use multiple query routers.** Use multiple mongos processes spread across multiple servers. A common deployment is to co-locate the mongos process on application servers, which allows for local communication between the application and the mongos process. The appropriate number of mongos processes will depend on the nature of the application and deployment.

## Application Patterns

MongoDB is an extremely flexible database due to its dynamic schema and rich query model. The system provides extensive secondary indexing capabilities to optimize query performance. Users should consider the flexibility and sophistication of the system in order to make the right trade-offs for their application. The following considerations will help you optimize your application patterns.

**Issue updates to only modify fields that have changed.** Rather than retrieving the entire document in your application, updating fields, then saving the document back to the database, instead issue the update to specific fields. This has the advantage of less network usage and reduced database overhead.

**Avoid negation in queries.** Like most database systems, MongoDB does not index the absence of values and negation conditions may require scanning all records in a result set. If negation is the only condition and it is not selective (for example, querying an orders table where 99% of the orders are complete to identify those that have not been fulfilled), all records will need to be scanned.

**Test every query in your application with `explain()`.** MongoDB provides the ability to view how a query will be evaluated in the system, including which indexes are used and whether the query is covered. This capability is similar to the Explain Plan and similar features in relational

databases. The [feedback from `explain\(\)`](#) will help you understand whether your query is performing optimally.

**Use covered queries when possible.** Covered queries return results from the indexes directly without accessing documents and are therefore very efficient. For a query to be covered all the fields included in the query must be present in the index, and all the fields returned by the query must be present in the index. To determine whether a query is a covered query, use the `explain()` method. If the `explain()` output displays `true` for the `indexOnly` field, the query is covered by an index, and MongoDB queries only that index to match the query and return the results.

**Avoid scatter-gather queries.** In sharded systems, queries that cannot be routed to a single shard must be broadcast to every shard for evaluation. Because these queries involve multiple shards for each request they do not scale as more shards are added.

**Your operational applications should only read from primaries.** Updates are typically replicated to secondaries quickly, depending on network latency. However, reads on the secondaries will not be consistent with reads on the primary. To increase read capacity in your operational system [consider sharding](#). Secondary reads can be useful for analytics and ETL applications as this approach will isolate traffic from operational workloads. You may read from secondaries if your application can tolerate eventual consistency.

**Use the most recent drivers from MongoDB.**

MongoDB supports drivers for nearly a dozen languages. These drivers are engineered by the same team that maintains the database kernel. Drivers are updated more frequently than the database, typically every two months. Always use the most recent version of the drivers when possible. Install native extensions if available for your language. Join the [MongoDB community mailing list](#) to keep track of updates.

**Ensure uniform distribution of shard keys.** When shard keys are not uniformly distributed for reads and writes, operations may be limited by the capacity of a single shard. When shard keys are uniformly distributed, no single shard will limit the capacity of the system.

**Use hash-based sharding when appropriate.** For applications that issue range-based queries, range-based

sharding is beneficial because operations can be routed to the fewest shards necessary, usually a single shard. However, range-based sharding requires a good understanding of your data and queries, which in some cases may not be practical. Hash-based sharding ensures a uniform distribution of reads and writes, but it does not provide efficient range-based operations.

## Schema Design & Indexes

MongoDB uses a binary document data model based called **BSON** that is based on the JSON standard. Unlike flat tables in a relational database, MongoDB's document data model is closely aligned to the objects used in modern programming languages, and in most cases it removes the need for complex transactions or joins due to the advantages of having related data for an entity or object contained within a single document, rather than spread across multiple tables. There are best practices for modeling data as documents, and the right approach will depend on the goals of your application. The following considerations will help you make the right choices in designing your schema and indexes for your application.

### **Store all data for a record in a single document.**

MongoDB provides ACID compliance at the document level. When data for a record is stored in a single document the entire record can be retrieved in a single seek operation, which is very efficient. In some cases it may not be practical to store all data in a single document, or it may negatively impact other operations. Make the trade-offs that are best for your application.

**Avoid large documents.** The maximum size for documents in MongoDB is 16 MB. In practice most documents are a few kilobytes or less. Consider documents more like rows in a table than the tables themselves. Rather than maintaining lists of records in a single document, instead make each record a document. For large media documents, such as video or images, consider using **GridFS**, a convention implemented by all the drivers that automatically stores the binary data across many smaller documents.

**Avoid unbounded document growth.** When a document is updated in the MongoDB MMAPv1 storage engine, the

data is updated in-place if there is sufficient space. If the size of the document is greater than the allocated space, then the document may need to be re-written in a new location. The process of moving documents and updating their associated indexes can be I/O-intensive and can unnecessarily impact performance. To anticipate future growth, the `usePowerOf2Sizes` attribute is enabled by default on each collection. This setting automatically configures MongoDB to round up allocation sizes to the powers of 2. This setting reduces the chances of increased disk I/O at the cost of using some additional storage.

An additional strategy is to manually pad the documents to provide sufficient space for document growth. If the application will add data to a document in a predictable fashion, the fields can be created in the document before the values are known in order to allocate the appropriate amount of space during document creation. Padding will minimize the relocation of documents and thereby minimize over-allocation. Learn more by reviewing the [record allocation strategies](#) in the documentation.

Avoiding unbounded document growth is a best practice schema design for any database, but the specific considerations above are **not relevant** to the WiredTiger storage engine which rewrites the document for each update.

**Avoid large indexed arrays.** Rather than storing a large array of items in an indexed field, instead store groups of values across multiple fields. Updates will be more efficient.

**Avoid unnecessarily long field names.** Field names are repeated across documents and consume space. By using smaller field names your data will consume less space, which allows for a larger number of documents to fit in RAM. Note that with its native compression, this is less of an issue for MongoDB databases configured with the WiredTiger storage engine.

**Use caution when considering indexes on low-cardinality fields.** Queries on fields with low cardinality can return large result sets. Avoid returning large result sets when possible. Compound indexes may include values with low cardinality, but the value of the combined fields should exhibit high cardinality.

**Eliminate unnecessary indexes.** Indexes are resource-intensive: even with compression enabled they

consume RAM, and as fields are updated their associated indexes must be maintained, incurring additional disk I/O overhead.

**Remove indexes that are prefixes of other indexes.**

Compound indexes can be used for queries on leading fields within an index. For example, a compound index on last name, first name can be used to filter queries that specify last name only. In this example an additional index on last name only is unnecessary; the compound index is sufficient for queries on last name as well as last name and first name.

**Avoid regular expressions that are not left anchored or rooted.** Indexes are ordered by value. Leading wildcards are inefficient and may result in full index scans. Trailing wildcards can be efficient if there are sufficient case-sensitive leading characters in the expression.

**Use index optimisations available in MongoDB**

**WiredTiger storage engine.** As discussed earlier, the WiredTiger engine compresses indexes by default. In addition, administrators have the flexibility to place indexes on their own separate volume, allowing for faster disk paging and lower contention.

## Disk I/O

While MongoDB performs all read and write operations through in-memory data structures, data is persisted to disk and the performance of the storage sub-system is a critical aspect of any system. Users should take care to use high-performance storage and to avoid networked storage when performance is a primary goal of the system. The following considerations will help you use the best storage configuration, including OS and filesystem settings.

**Readahead size should be set to 32.** Set readahead to 32 or the size of most documents, whichever is larger. Most disk I/O in MongoDB is random. If the readahead size is much larger than the size of the data requested, a larger block will be read from disk. This has two undesirable consequences: 1) the size of the read will consume RAM unnecessarily, and 2) more time will be spent reading data than is necessary. Both will negatively affect the performance of your MongoDB system. Readahead should not be set lower than 32.

**Use EXT4 or XFS file systems; avoid EXT3.** EXT3 is quite old and is not optimal for most database workloads. For example, MongoDB pre-allocates space for data. In EXT3 pre-allocation will actually write 0s to the disk to allocate the space, which is time consuming. In EXT4 and XFS pre-allocation is performed as a logical operation, which is much more efficient.

**Disable access time settings.** Most file systems will maintain metadata for the last time a file was accessed. While this may be useful for some applications, in a database it means that the file system will issue a write every time the database accesses a page, which will negatively impact the performance and throughput of the system.

**Don't use hugepages.** Do not use hugepages virtual memory pages, MongoDB performs better with normal virtual memory pages.

**Use RAID10.** The performance of RAID5 is not ideal for MongoDB. RAID0 performs well but does not provide sufficient fault tolerance. RAID10 provides the best balance between performance and fault tolerance. Place journal, log and data files on different devices. By using separate storage devices for the journal and data files you can increase the overall throughput of the disk subsystem. Because the disk I/O of the journal files tends to be sequential, SSD may not provide a substantial improvement and standard spinning disks may be more cost effective.

## Considerations for Amazon EC2

Amazon EC2 is an extremely popular environment for MongoDB deployments. MongoDB has worked with Amazon to determine best practices in order to ensure users have a great experience with [MongoDB on EC2](#).

**Deploy MongoDB with MMS.** The [MongoDB Management Service \(MMS\)](#) makes it easy for you to provision, monitor, backup and scale MongoDB. MMS provisions your AWS virtual machines with an optimal configuration for MongoDB, including configuration of the readahead and ulimits. You can choose between different instance types, Linux AMIs, EBS volumes, ephemeral storage and regions. Provisioned IOPS and EBS-optimized



instances provide substantially better performance for MongoDB systems. Both can be configured from MMS.

## Considerations for Benchmarks

Generic benchmarks can be misleading and misrepresentative of a technology and how well it will perform for a given application. MongoDB instead recommends that users model and benchmark their applications using data, queries, hardware, and other aspects of the system that are representative of their intended application. The following considerations will help you develop benchmarks that are meaningful for your application.

**Model your benchmark on your application.** The queries, data, system configurations, and performance goals you test in a benchmark exercise should reflect the goals of your production system. Testing assumptions that do not reflect your production system is likely to produce misleading results.

**Create chunks before loading, or use hash-based sharding.** If range queries are part of your benchmark use range-based sharding and [create chunks before loading](#). Without pre-splitting, data may be loaded into a shard then moved to a different shard as the load progresses. By pre-splitting the data documents will be loaded in parallel into the appropriate shards. If your benchmark does not include range queries, you can use hash-based sharding to ensure a uniform distribution of writes.

**Disable the balancer for bulk loading.** Prevent the balancer from rebalancing unnecessarily during bulk loads to improve performance.

**Prime the system for several minutes.** In a production MongoDB system the working set should fit in RAM, and all reads and writes will be executed against RAM. MongoDB must first page the working set into RAM, so prime the system with representative queries for several minutes before running the tests to get an accurate sense for how MongoDB will perform in production.

**Monitor everything to locate your bottlenecks.** It is important to understand the bottleneck for a benchmark. Depending on many factors any component of the overall

system could be the limiting factor. A variety of popular tools can be used with MongoDB – [many are listed in the manual](#).

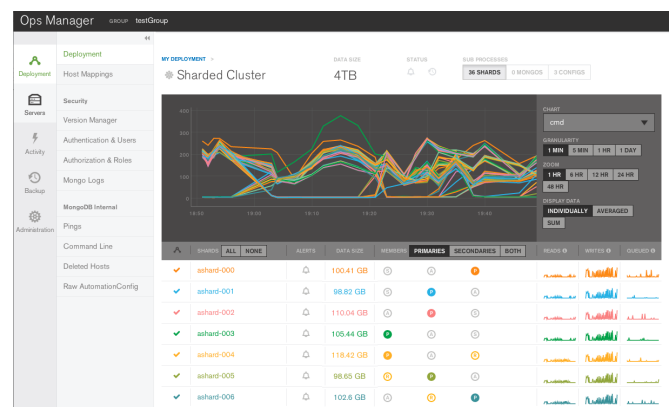
The most comprehensive tools for monitoring MongoDB are Ops Manager, available as a part of [MongoDB Enterprise Advanced](#). Featuring charts, custom dashboards, and automated alerting, Ops Manager tracks 100+ key database and systems metrics including operations counters, memory and CPU utilization, replication status, open connections, queues and any node status. The metrics are securely reported to Ops Manager where they are processed, aggregated, alerted and visualized in a browser, letting administrators easily determine the health of MongoDB in real-time. The benefits of Ops Manager are also available in the SaaS-based MMS (mentioned earlier), hosted by MongoDB in the cloud.

In addition to monitoring, Ops Manager and MMS provide automated deployment, upgrades, and backup.

### Use mongoperf to characterize your storage system.

[mongoperf](#) is a free, open-source tool that allows users to simulate direct disk I/O as well as memory mapped I/O, with configurable options for number of threads, size of documents and other factors. This tool can help you to understand what sort of throughput is possible with your system, for disk-bound I/O as well as memory-mapped I/O. The MongoDB Blog provides a good overview of [how to use the tool](#).

**Follow configuration best practices.** Review the [MongoDB production notes](#) for the latest guidance on



**Figure 1:** Ops Manager & MMS provides real time visibility into MongoDB performance.



packages, hardware, networking and operating system tuning.

## We Can Help

We are the MongoDB experts. Over 2,000 organizations rely on our commercial products, including startups and more than a third of the Fortune 100. We offer software and services to make your life easier:

**MongoDB Enterprise Advanced** is the best way to run MongoDB in your data center. It's a finely-tuned package of advanced software, support, certifications, and other services designed for the way you do business.

**MongoDB Management Service (MMS)** is the easiest way to run MongoDB in the cloud. It makes MongoDB the system you worry about the least and like managing the most.

**Production Support** helps keep your system up and running and gives you peace of mind. MongoDB engineers help you with production issues and any aspect of your project.

**Development Support** helps you get up and running quickly. It gives you a complete package of software and services for the early stages of your project.

**MongoDB Consulting** packages get you to production faster, help you tune performance in production, help you scale, and free you up to focus on your next release.

**MongoDB Training** helps you become a MongoDB expert, from design to operating mission-critical systems at scale. Whether you're a developer, DBA, or architect, we can make you better at MongoDB.

## Resources

For more information, please visit [mongodb.com](http://mongodb.com) or contact us at [sales@mongodb.com](mailto:sales@mongodb.com).

Case Studies ([mongodb.com/customers](http://mongodb.com/customers))

Presentations ([mongodb.com/presentations](http://mongodb.com/presentations))

Free Online Training ([university.mongodb.com](http://university.mongodb.com))

Webinars and Events ([mongodb.com/events](http://mongodb.com/events))

Documentation ([docs.mongodb.org](http://docs.mongodb.org))

MongoDB Enterprise Download ([mongodb.com/download](http://mongodb.com/download))

