What's New in MongoDB 2.6

September 2014

Table of Contents

Introduction

MongoDB 2.6 is the newest and most significant release of the world's fastest growing database. It builds on five years of innovation and hundreds of thousands of deployments to provide a new foundation for the database, drivers and sophisticated management tools that make operating MongoDB simple at any scale

MongoDB 2.6 enables new classes of use cases, non-stop operations and enterprise-grade security. Key features of the release include:

- **Index Intersection.** Makes it easier to run ad-hoc analyses to answer evolving business questions. Developers no longer need to predict all data access patterns in advance as more than one index can be used to satisfy a query.
- **Improved Scalability & Performance.** Provides more efficient use of network resources; oplog processing is 75% faster; classes of scan, sort, $in and $all performance are significantly improved; and bulk operators for writes improve updates by as much as 5x.
- **Pipelined Data Transformations.** Makes it possible to perform multi-step data enrichment and transformations natively in the database using a simple declarative interface. With the new $out stage, result sets from the aggregation pipeline can be written to a named collection with no limit to the output size, subject to the underlying storage system.
- **Text Search.** Search can be delivered as a feature of their application without the added complexity of deploying a dedicated search engine. Now fully integrated into query language and the Aggregation Framework, Text Search provides rich, powerful search abilities for 15 languages.
- **Enterprise Security.** Building on the existing MongoDB security features including Kerberos authentication and SSL encryption, MongoDB 2.6 adds collection-level authorization, user- defined roles and field-level redaction. **MongoDB Enterprise** includes support for auditing, LDAP integration and x509 authentication.
- **Bulk Update Operators.** A single write concern can be specified for an entire write operation, making it much simpler and more efficient to load large batches

of data to MongoDB. Bulk operations automatically parallelize updates across the system, returning a report of failed operations that can be retried by the application.

- **Non-Stop Operations.** Indexing auto-resumes after restart, and can be performed in the background, yielding to foreground operations; MaxTimeMS allows operators and developers to specify auto-cancellation of queries on a per-operation basis, providing better control of resource utilization; mixed SSL connections; expanded SNMP support; more efficient repair operations; and a new default space allocation configuration provides more predictable performance.
- **Expanded Management Tools.** MongoDB Management Service (MMS) is a cloud service for managing MongoDB, created by the engineers who develop the database. It is the easiest, most reliable way to operate MongoDB at scale. MMS provides automated provisioning and management, proactive monitoring and continuous, incremental backup and point-in-time recovery.

You can learn more about each of these areas in this Guide.

Real Time Analytics & BI: Index Intersection & Transformations

The emergence of new data sources such as social media, mobile applications and sensor-equipped "Internet of Things" networks is enabling organizations to extend BI into every area of their business. With the right technologies, users can unlock real-time insight and discovery into such areas as operational performance, customer satisfaction and competitor behavior.

With its rich document model and powerful analytical capabilities over high volumes of structured, semi-structured and unstructured data, MongoDB provides a foundation to evolve BI to support real-time analytics for big data applications. These capabilities are enhanced in MongoDB 2.6 with index intersection, pipelined data transformations, new operators and the broadest integration with leading BI and analytics tools.

These enhancements extend MongoDB's lead with the best BI & analytics capabilities of any NoSQL database. Organizations can explore, analyze and monetize multi-structured data with the fastest speed to insight and greatest ease-of-use.

Index Intersection

New in MongoDB 2.6, index intersection enables the query planner to use more than one index to resolve a query. With index intersection developers can address more queries with simple indexes and with less upfront index design.

As an example, consider a sales reporting application: A product manager wants to identify all customers who have ordered more than a given quantity of a specific part number. Using index intersection the existing indexes for part number and quantity can be combined (intersected) to optimize the query, rather than requiring a separate compound index. This also results in reduced overhead to the working set size, and more efficient updates.

Index intersection currently supports the intersection of two indexes and is best used when the cardinality of the result sets are roughly equivalent, and especially for those queries that can be resolved from covered indexes. In cases where multi-field predicates are known in advance, queries can be resolved more quickly with a compound index.

More information on index **intersection is in the documentation**.

Pipelined Data Transformation

MongoDB now makes it possible to perform multi-step data enrichment and transformations natively in the database using a simple declarative interface. Pipelined data transformations enable such applications as lightweight ETL within MongoDB, eliminating the time and cost of moving data between multiple components of an analytics pipeline.

With the new `$out stage`, result sets from the aggregation pipeline can be written to a named collection with no limit to the output size (subject to the underlying storage system). Existing collections are atomically replaced with new results, while maintaining previously

defined indexes to ensure queries can always be returned efficiently.

As part of implementing the new **`$out stage`**, the `db.collection.aggregate` method can now return a cursor with result sets of any size. Previous versions returned all results in a single document, and so the result set was subject to a size limit of 16 megabytes.

New Update Operators

MongoDB's new **`$mul`**, **`$min`** and **`$max`** operators enable developers to efficiently manipulate numeric values with less code. The **`$currentDate`** operator automati- cally sets the value of a field to the current date.

New modifiers for the **`$push`** operator simplify the manipulation of data within arrays, including the sorting, slicing and positioning of array elements.

The documentation includes more detail on the **new update operators**.

New Set Operators

New Set operators in MongoDB 2.6 enable richer $project processing and transformations of arrays, including equalities, unions and intersections.

Refer to the documentation for a description of each of these **new set operators**.

BI Tools Integration

To make online big data actionable it needs to be accessible to BI and analytics tools. Whether for reporting and visualization, or for mash-ups with other operational data sources, MongoDB offers integration with a broader set of the leading tools than any other NoSQL or online big data technology, including:

- Actuate
- Alteryx
- Informatica
- Jaspersoft
- Logi Analytics
- MicroStrategy
- Pentaho
- Qliktech
- SAP Lumira
- Talend

These integrations open up three key use-cases for MongoDB within a BI and analytics platform:

- As a **conventional data source** for regular ETL processes integrating data into the Enterprise Data Warehouse.
- As a **"single view aggregration,"** replicating and consolidating data from operational and EDW sources, allowing for cross-function, 360-degree view reporting and visualization.
- As a **data store** enabling real-time analytics and dashboards to be generated against live, operational data.

You can learn more by downloading the MongoDB whitepaper: **Bringing Online Big Data to Business Intelligence & Analytics**

Integrated Search

Search has become one of the primary means of accessing information in applications. With the popularity of Internet search engines, many users expect search to be delivered as a first class feature. Using MongoDB's native text search functionality, developers can deliver search as a feature of their application without the added complexity of using dedicated search engines. Analysts can also unlock text-aware analytics and insight from semi-structured and unstructured data in real-time.

Moving out of beta, text search is now production-ready, and offers new functionality, including:

- Integration with MongoDB's query engine. Text search can be combined with general query operators to

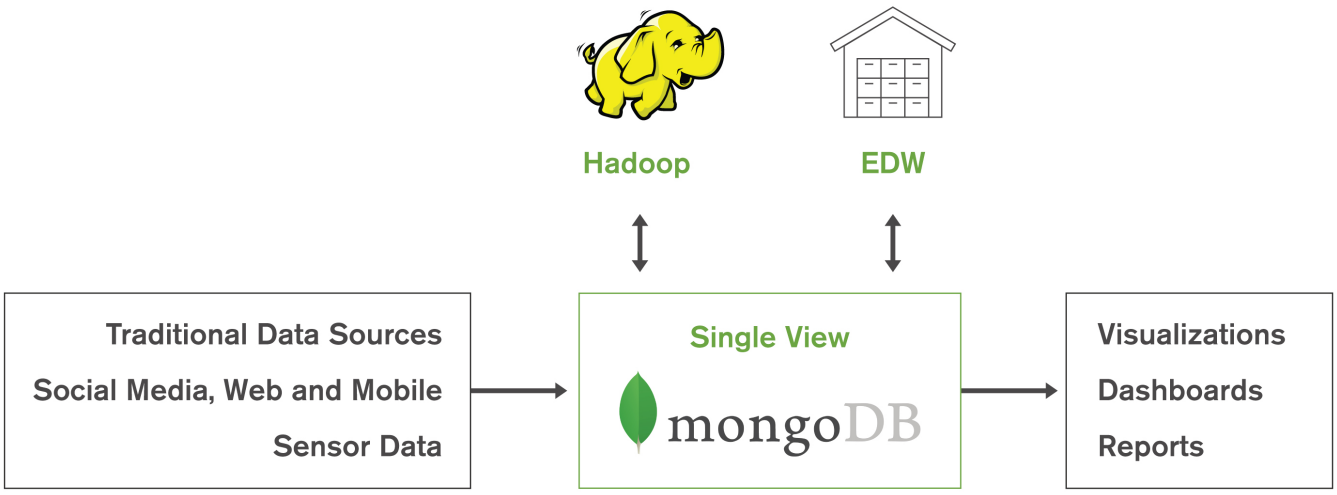

Figure 1: Single View of the Business, Powered by MongoDB

provide richer queries with the ability to limit, skip, sort and filter results. For example, a user could search a collection of blog posts for certain phrases, but limit the search to posts from the last seven days using an additional condition.

- Multi-language document support.
- Text search expressions can be used in the Aggregation Framework, providing deeper analytics with counting and grouping of text matches.

MongoDB supports core text search features that meet the needs of many applications, including:

- Relevance ranking.
- Boolean operators.
- Language-specific tokenization and stemming.
- Fielded search.
- Field-weighted scoring.
- Stop words.
- Type-aware indexes.
- Wildcards.

To ensure consistency, updates to data also update text indexes atomically in real-time, enabling users to combine text search with the ease-of-use, scalability and High Availability (HA) of MongoDB.

Reducing the Cost of Search

Unlike many other big data technologies or RDBMS, users can run text search queries in-place over MongoDB data, thereby reducing the cost and operational overhead to deliver search functionality. Integrating text search in MongoDB avoids the need to provision an external search cluster and replicate data to it for common search functions.

- **Up to Date.** Storage and analysis of real time, operational data in the database, rather than potentially stale data copied to an external cluster.
- **Easy to Manage.** Operational simplicity, avoiding the complexity of spinning up a dedicated search cluster with its own management, configuration, scaling and HA concerns.
- **Consistent Experience.** Simpler development, as developers only need to work with the MongoDB API to query their data.

You can learn more in the **MongoDB text search tutorial**.

Enterprise Security

With increased regulatory compliance, heightened concerns around privacy and growing risk from hackers and organized crime, the need to secure access to data has never been more urgent. Industry research indicates

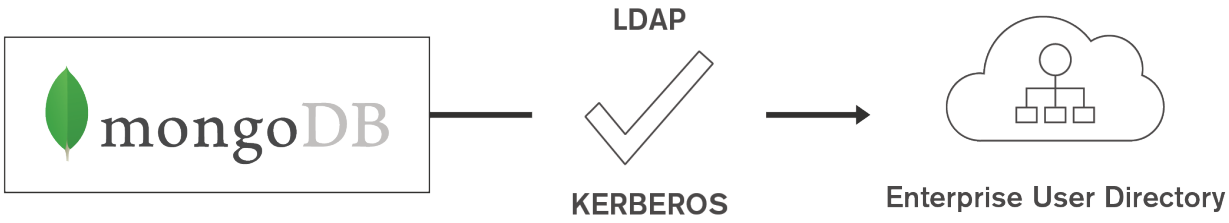

Figure 2: Integrating MongoDB with Centralized User Access Controls

that 96% of all data thefts come from records stored within databases.[1]

Data collected from social media, mobile devices and sensor networks has become as sensitive as traditional transactional data generated from back-office systems. For this reason, big data technologies must evolve to meet the regulatory compliance standards demanded by industry and government. While these standards address different industries and geographies, they have four common elements:

- Restricting access to data, enforced via predefined privileges and security levels.
- Measures to protect against the accidental or malicious disclosure, loss, destruction or damage of sensitive data.
- The separation of duties when running applications and accessing data.
- Recording the activities of users, administrative staff and applications in accessing and processing data.

These requirements inform MongoDB's security architecture. Building on the support in MongoDB 2.4 for roles, Kerberos authentication and SSL encryption, the latest 2.6 release adds additional authentication integration, authorization and auditing capabilities. Each of the enhancements are discussed below, and you can learn more by reviewing the **security documentation**.

Authentication With LDAP & X.509 Certificates

Authentication is designed to confirm the identity of entities accessing the database. MongoDB 2.4 supports authentication both locally within the database and via integration with external mechanisms using Kerberos. MongoDB 2.6 adds support for LDAP and x.509 certificates.

LDAP in MongoDB 2.6

Many organizations use Lightweight Directory Access Protocol (LDAP) to standardize and simplify the way large numbers of users are managed across internal systems and applications. In many cases, LDAP is also used as the centralized authority for user access control to ensure that internal security policies comply with corporate and regulatory guidelines.

With LDAP integration, MongoDB can authenticate users directly against corporate LDAP infrastructure, eliminating the need to duplicate password management between LDAP directories and MongoDB's internal authentication controls. Note that MongoDB currently supports LDAP authentication, and not authorization. See the following section of the whitepaper to learn more about the authorization controls available in MongoDB.

Administrators **can configure MongoDB** to authenticate users via Linux PAM or by proxying authentication requests to a specified LDAP service.

LDAP integration is available with **MongoDB Enterprise**, more details of which are included later in the whitepaper.

x.509 Certificates in MongoDB 2.6

With support for x.509 certificates, MongoDB can be integrated with existing information security infrastructure and certificate authorities, supporting both user and inter-node authentication.

1. http://www.verizonenterprise.com/resources/reports/rp*data-breach-investigations-report-2012-ebk*en_xg.pdf

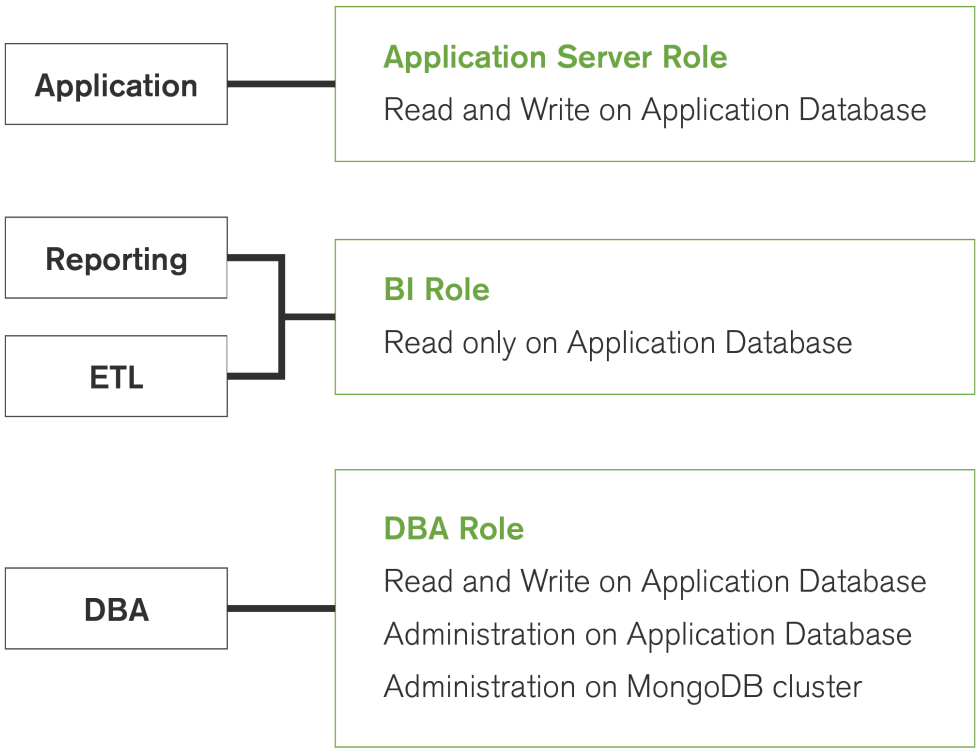

Figure 3: MongoDB User Defined Roles Permit Separations of Duty

Users can be be authenticated to MongoDB using client certificates rather than self-maintained, potentially vulnerable passwords.

Inter-cluster authentication and communication between MongoDB replica set and sharded cluster nodes can be secured with x.509 certificates rather than keyfiles, ensuring stricter membership controls with less administrative overhead (i.e., by eliminating the shared password used by keyfiles.)

Instructions for configuration are described in the **MongoDB and x.509 certificates tutorial**. x.509 certificate authentication is used with a secure SSL connection. To use SSL, users must either be running MongoDB Enterprise or build MongoDB locally using **SCons with the --ssl option**.

Authorization With User-Defined Roles & Field Level Redaction

Once an entity has been authenticated, authorization governs what that entity is entitled to do in the database. Privileges are assigned to user roles, which define a specific set of actions that users can perform against the database. MongoDB 2.6 introduces user-defined roles and field level redaction.

User-Defined Roles

MongoDB 2.4 provides the ability to differentiate between user and administrator privileges with built-in roles. MongoDB 2.6 extends authorization capabilities with User-Defined Roles, enabling administrators to assign finely-grained privileges to users and applications, based on the specific functionality they require. MongoDB provides the ability to specify user privileges at both the database and collection levels.

Privileges are assigned to roles, and roles are in turn assigned to users. For example:

- Classes of users and applications can be assigned privileges to insert data, but not to update or delete data from the database. BI and analytics applications can be given rights to read data, but never update it.
- DBAs may be assigned privileges that enable them to create collections and indexes on the database, while developers are restricted to CRUD operations.

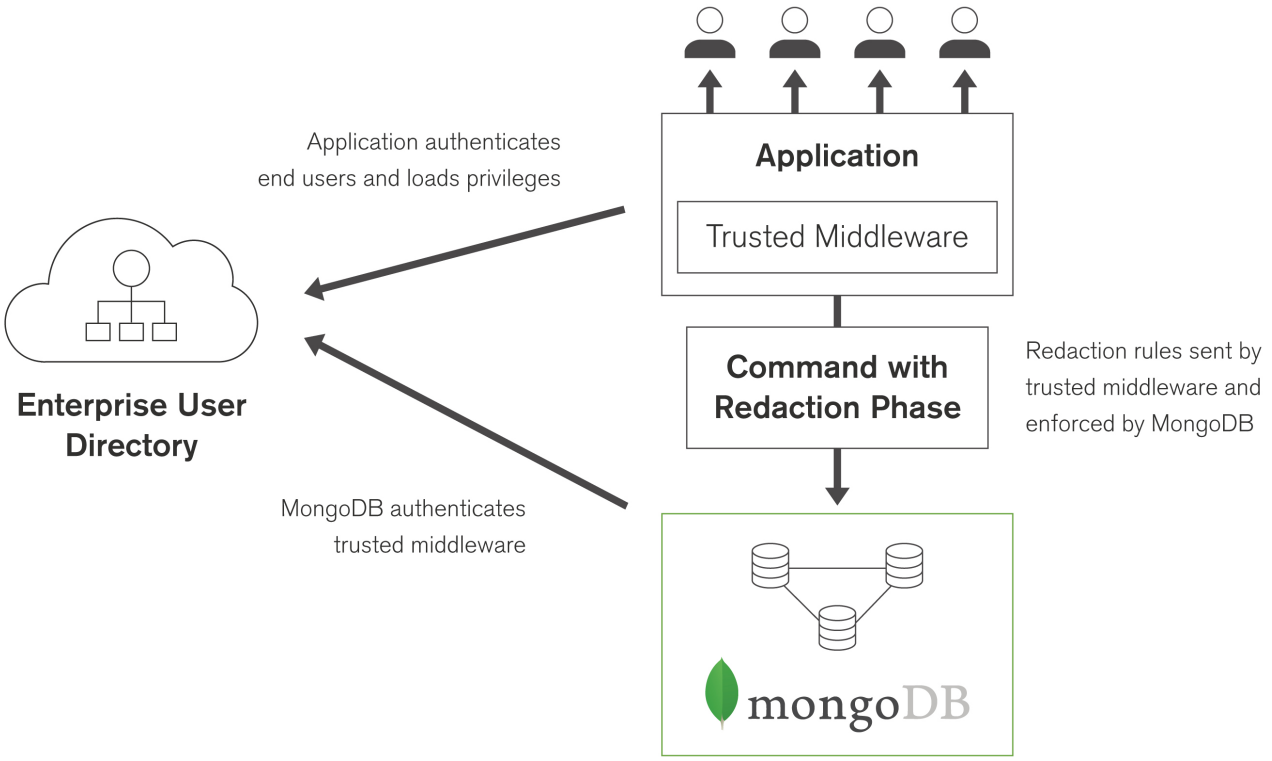

Figure 4: MongoDB Field Level Redaction Restricts Access to Sensitive Data

- Certain administrator roles may have cluster- wide privileges to build replica sets and configure sharding, while others are restricted to creating new users or inspecting logs.
- Processes for monitoring MongoDB clusters can be restricted to run just those commands that retrieve server status, without having full administrative access to perform database operations.
- Within a multi-tenant environment, 'landlord' developers and administrators can be assigned permissions across physical databases, while 'tenant' developers and administrators can be granted a more limited set of actions across logical databases or individual collections. This functionality enables a clear separation of duties and control, both between and within organizations.

To ensure ease of account provisioning and maintenance, roles can be delegated across teams, ensuring the enforcement of consistent policies across specific functions within the organization. Review the **authorization section** of the documentation to learn more about roles in MongoDB. When combined with the auditing capabilities available in MongoDB Enterprise, customers can now define specific administrative actions per role, and then log all of those actions. As a result, the organization is able to enforce end-to-end operational control and maintain insight of actions for compliance and reporting.

Field Level Redaction

MongoDB's field level redaction allows building field level access control in trusted middleware. MongoDB's **Aggregation Pipeline** includes a new **redaction stage**, providing a method to restrict the content of a returned document on a per-field level. Permissions can be based on both the content of the document and on specific user privileges, based on security labels. Access control policies can be described using the MongoDB query language, making it simple for developers to implement the required controls.

Since data is redacted before it is returned to the application, exposure of sensitive information is eliminated. Field level redaction is applicable to a wide range of sensitive data including personally identifiable information such as names, addresses, social security numbers, birthdates and bank account numbers.

As the application must pass the redaction logic to the database on each request, it relies on trusted middleware running in the application to ensure the redaction pipeline stage is appended to any query that requires the redaction logic.

Auditing

MongoDB Enterprise logs all administrative actions made against the database. Schema operations (such as creating or dropping databases, collections and indexes), replica set reconfigurations along with authentication and authorization activities are all captured, along with the administrator's identity and timestamp of the operation, enabling compliance and security analysis.

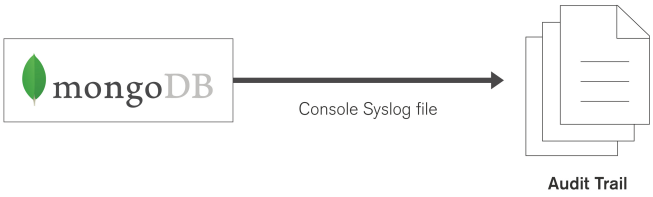

Figure 5: MongoDB Maintains an Audit Trail of Administrative Actions Against the Database

By default, MongoDB auditing logs all administrative actions, but can also be configured with filters to capture only specific events. The audit log can be written to multiple destinations in a variety of formats including to the console and syslog (in JSON format),and to a file (JSON or BSON), which can then be loaded to MongoDB and analyzed to find relevant events. Each MongoDB server logs events to its local destination. The DBA can then use their own tools to merge these into a single log, enabling a cluster-wide view of operations that affected multiple nodes.

The **MongoDB auditing documentation** includes information on how to configure auditing and all of the operations that can be captured. To support auditing of read and write activity to the database, MongoDB has also been certified with IBM's InfoSphere Guardium. The **IBM DeveloperWorks zone** provides an overview of the integration and a configuration guide.

Non-Stop Operations

MongoDB 2.6 introduces a range of enhancements to improve performance and scalability while reducing operational overhead. This helps organizations improve SLAs and reduce TCO.

Improving Performance & Scalability

Query Router Connection Pooling

Connection pools for the MongoDB query routers can now be re-used by multiple MongoDB servers. This reduces the total number of connections needed in highly concurrent systems, lowering resource consumption by up to 5x. As a result, users can scale out to larger clusters supporting higher user load with less hardware.

Resource Overload Protection

Operations staff can maintain SLAs by ensuring database resources are protected from over- consumption by rogue operations. The new `$maxTimeMS` option enables administrators to abort MongoDB queries and commands automatically if they exceed a pre-configured maximum time limit, with error messages returned to the client.

Administrators can coordinate time-outs between the application, the driver and the database to prevent resource overload and ensure predictable performance. Time-outs can also be configured per query, affording high user flexibility.

Reducing Operational Overhead

Background Secondary Indexing

Index builds running in the foreground of a MongoDB instance will block other database operations.

To provide continuous availability, MongoDB therefore supports the creation of indexes in the background when they are running on the primary member of a replica set.

MongoDB 2.6 extends this capability to **secondary members of the replica set**, which now replicate and perform the index build in the background as well.

As a result, both primary and secondary replicas remain fully online, without operators having to manage rolling restarts. It is important to note that any background index build operation is computationally expensive and will reduce the throughput of the MongoDB server while in progress. In addition, index creation operations that are inter- rupted due to a failure can now automatically resume when the node recovers.

Mixed SSL Connections

Users now have the flexibility to run SSL and non-SSL connections to the same MongoDB server. As a result, administrators can avoid MongoDB downtime as they implement SSL across the cluster.

You can learn more by reading the **tutorial** on upgrading a MongoDB cluster to SSL.

Bulk Write Operations

It is now possible to configure a single write concern for an entire bulk write operation, making it much simpler to load large batches of data to MongoDB. Bulk loads can be configured to either:

1. Complete the loading operation, reporting any write errors for later retries;
2. Stop the loading operation immediately if an error is returned.

You can read more about **bulk write operations** in the documentation.

	MongoDB Community Edition	MongoDB Enterprise
Index Intersection	✓	✓
Pipelined Data Transformation	✓	✓
New Update & Set Operators	✓	✓
Search	✓	✓
User-Defined Roles	✓	✓
Field Level Redaction	✓	✓
Query Router Connection Pooling	✓	✓
Bulk Write Operations	✓	✓
Resource Overload Protection	✓	✓
Background Secondary Indexing	✓	✓
Mixed SSL Connections		✓
x.509 Certificates		✓
LDAP Authentication		✓
Auditing		✓
Expanded SNMP		✓

Table 1: Availability of Features in MongoDB 2.6

Expanded SNMP Support

MongoDB Enterprise offers SNMP support, which can be used to integrate MongoDB with external monitoring solutions.

Using MongoDB 2.6, customers now have access to nearly the full range of metrics provided by the `db.serverStatus() command`, providing more detailed server metrics.

You can learn more by reading the **tutorial** on configuring SNMP for MongoDB.

MongoDB Management Service (MMS)

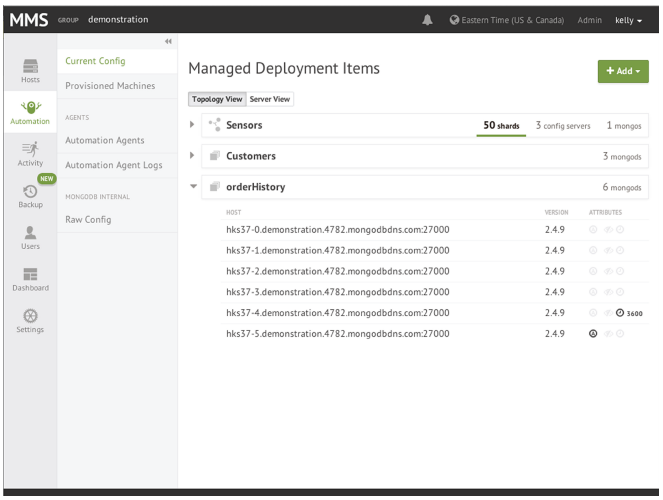

Figure 6: MMS Automation - Simple and Sophisticated UI

MMS is a cloud service for managing MongoDB, created by the engineers who develop the database. It is the easiest, most reliable way to operate MongoDB at scale. Available as a managed service in the cloud, MMS provides an integrated suite of applications that manage the complete lifecycle of the database:

- Automated provisioning and management with a single click and zero-downtime upgrades;
- Proactive monitoring with visibility into current and historic performance of MongoDB, and automated alerting on 100+ system metrics;
- Disaster recovery with continuous, incremental backup and point-in-time recovery.

Getting Started with MongoDB 2.6

The latest MongoDB production release is available from the **downloads page**, along with development releases previewing future functionality.

As already mentioned, you can also **download MongoDB Enterprise** free of charge for development.

Review the **release notes** on upgrading from previous releases to MongoDB 2.6. The documentation includes an upgrade checklist and provides instructions on upgrading from a standalone MongoDB instance, a replica set and a sharded cluster.

What We Sell

We are the MongoDB experts. Over 1,000 organizations rely on our commercial products, including startups and more than 30 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.

Contact us to learn more, or visit www.mongodb.com.

Resources

For more information, please visit mongodb.com or contact us at sales@mongodb.com.

Case Studies (mongodb.com/customers)
Presentations (mongodb.com/presentations)
Free Online Training (university.mongodb.com)
Webinars and Events (mongodb.com/events)
Documentation (docs.mongodb.org)
MongoDB Enterprise Download (mongodb.com/download)

mongoDB

New York • Palo Alto • Washington, D.C. • London • Dublin • Barcelona • Sydney • Tel Aviv
US 866-237-8815 • INTL +1-650-440-4474 • info@mongodb.com