Technical Report

End-to-End Storage Provisioning for MongoDB
Deploying NetApp OnCommand Workflow Automation for MongoDB

John Elliott, NetApp
April 2018 | TR-4674

Abstract

This technical report explains the benefits of using NetApp® OnCommand® Workflow Automation for MongoDB storage deployment. The report covers the workflow processes for MongoDB replica sets and sharded clusters. It also supports MongoDB deployments that span multiple data centers.

n NetApp®

TABLE OF CONTENTS

LIST OF FIGURES

1 Big Data Challenges

Effectively leveraging the massive and growing amounts of big data is mission critical for enterprise organizations. Enterprises must not only monitor and protect big data, they must also quickly extract valuable business intelligence from huge amounts of unstructured data. In the current highly competitive environment, that intelligence is a requirement for survival and profitability.

Features that were once optional, such as robust, storage-efficient data protection and instant, free clones of data for DevOps, are now required. Because of the large size of these datasets, cost savings are more important than ever. In many cases, the need to reduce costs drives enterprise-class organizations to the cloud for data protection and for workflow execution. Given the mission-critical nature of these workflows, continuous availability of both the data and the applications is necessary. Downtime, slow performance, and data loss are completely unacceptable.

In addition to the requirements for enterprise data protection and management features, enterprise organizations are under increasing pressure to simplify and decrease the time to roll out new deployments. This business pressure is driven by the need to reduce costs, increase the agility of the organization to quickly monetize new data, and leverage the data to improve customer service and gain market share.

This technical report focuses on MongoDB powered by NetApp ONTAP® data management software deployed using NetApp OnCommand Workflow Automation (WFA). MongoDB is the NoSQL database that is most widely used by enterprise organizations today. ONTAP software brings enterprise-class data protection and data management to MongoDB. MongoDB professionals can use WFA to deploy MongoDB with ONTAP even if they have limited knowledge of storage because WFA uses the same MongoDB vocabulary and concepts. With WFA, best practices are built in and storage deployment is automated, eliminating costly deployment errors and greatly reducing deployment time. WFA supports simple on-premises data center deployments, and it also supports and simplifies more complex configurations that span multiple data centers.

This report begins with a brief discussion of each of the key technologies required for deployment and a description of the solution. Section 3 describes a few high-value use cases, and section 4 provides basic instruction on how to implement WFA with MongoDB powered by ONTAP.

2 Solution

2.1 MongoDB

MongoDB is a very popular open-source scale-out NoSQL database. It powers modern big data analytics applications that require low latency for reads and writes, high availability, and advanced data management. Key use cases for MongoDB include real-time analytics, product catalogs, content management, and mobile applications.

2.2 NetApp Data Fabric

Data Fabric is the NetApp vision for the future of data management. A Data Fabric seamlessly connects different data management environments across disparate clouds into a cohesive, integrated whole. The NetApp Data Fabric helps organizations maintain control and choose the way they manage, secure, protect, and access their data across the hybrid cloud, no matter where it is. Although a Data Fabric evolves constantly, organizations can start taking advantage of it today by using NetApp technologies that enable data management and seamless data movement across the hybrid cloud. For more information about the Data Fabric powered by NetApp, see WP-7128: NetApp Data Fabric Architecture Fundamentals: Building a Data Fabric Today.

2.3 NetApp OnCommand Workflow Automation for MongoDB

The NetApp OnCommand WFA for MongoDB solution uses the MongoDB vocabulary to provision storage for any MongoDB deployment. The automated approach enables the storage administrator to delegate the provisioning of new MongoDB storage configurations to the database administrator (DBA), both within a single data center and across multiple data centers.

> **NetApp Workflow Automation for MongoDB**
>
> The NetApp WFA for MongoDB is the industry's only multiple-data-center, end-to-end provisioning automation solution. It enables flexible deployment of MongoDB with Data Fabric for on-premises, hybrid cloud, and multicloud configurations.

NetApp WFA for MongoDB simplifies any deployment of NetApp ONTAP data management software for MongoDB by allowing a DBA or developer to provision storage in terms of database architecture and dataset size. It automatically implements best practices and simplifies provisioning for complex environments, such as multiple-data-center deployments. In addition, it fully supports both replica sets and sharded clusters. As a result, organizations that use MongoDB can quickly respond to the changing needs of their customers. MongoDB environments can be configured much more quickly, which reduces the time required for the enterprise to begin monetizing data.

The benefits of using WFA for MongoDB include:

- Standardization of manual processes
- Reduction in errors
- Maximization of (high) availability
- Improved QoS
- Self-service for DBAs and developers
- Built-in adherence to best practices
- Freeing up valuable bandwidth so that teams can focus on more strategic areas
- Decreased application time to market
- Access to the industry's only multiple-data-center, end-to-end provisioning automation solution

WFA Template for MongoDB Deployment

The WFA solution provides a template that guides the user through the entire process of provisioning and configuring storage for a new MongoDB environment. The DBA specifies the WFA storage sizing and configuration by using familiar MongoDB vocabulary. Simply put, the DBA provides a description of the required MongoDB environment, and the WFA automatically configures the required storage.

The user does not have to specify the volumes, LUNs, igroups, or virtual interfaces (VIFs). The workflow template converts the MongoDB architectural requirements into storage requirements, and then provisions that storage. Both MongoDB and NetApp ONTAP storage best practices are built in, resulting in optimal performance and eliminating the potential for costly configuration errors.

Users can download WFA for MongoDB and the corresponding product documentation from http://mysupport.netapp.com/.

2.4 NetApp and MongoDB Solution with WFA

Figure 1 shows a basic MongoDB deployment with ONTAP storage. The on-premises instance has been provisioned by using WFA. Once that instance has been created, NetApp SnapMirror® replication technology can be used to replicate data copies into a public or private cloud. In this example, data is mirrored to an ONTAP cloud instance in the public cloud or to an ONTAP NetApp Private Storage (NPS)

system, in a colocated data center. Data copied to either location can be accessed by public cloud compute resources, and public cloud storage can be used as a low-cost data archive. For the on-premises deployment, WFA supports both replica sets and sharded clusters.

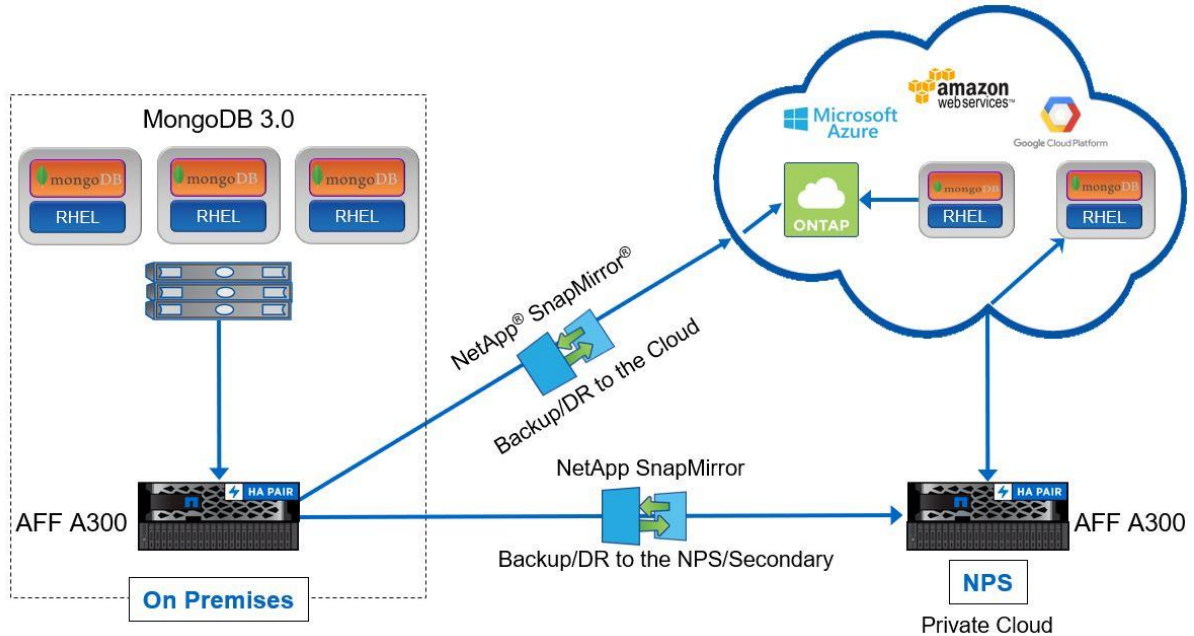**Figure 1) Simple on-premises configuration with NetApp Data Fabric.**

Figure 2 shows MongoDB deployed across three data centers. This type of deployment is increasingly popular with enterprise organizations to guarantee 100% data availability, even in the event of the catastrophic failure of an entire data center. In this configuration, an entire data center can go offline and the data will still be available. NetApp is the industry leader in providing automation and simplified deployment for this type of configuration.

Figure 2 depicts three data centers and a MongoDB sharded cluster configured with three replica sets. Each replica set contains:

- 1 primary replica
- 4 secondary replicas
- 1 arbiter, which has voting rights for replica failover, but contains no data

The replicas are spread across the three data centers in such a way that the loss of an entire data center does not impact data availability. The data in each data center is stored in an ONTAP storage system and uses NetApp SnapCenter® software for data protection and copy management. The entire ONTAP portfolio of tools and software is available in each data center, including data mobility across the NetApp Data Fabric.

Figure 2) Triple data center design with WFA.
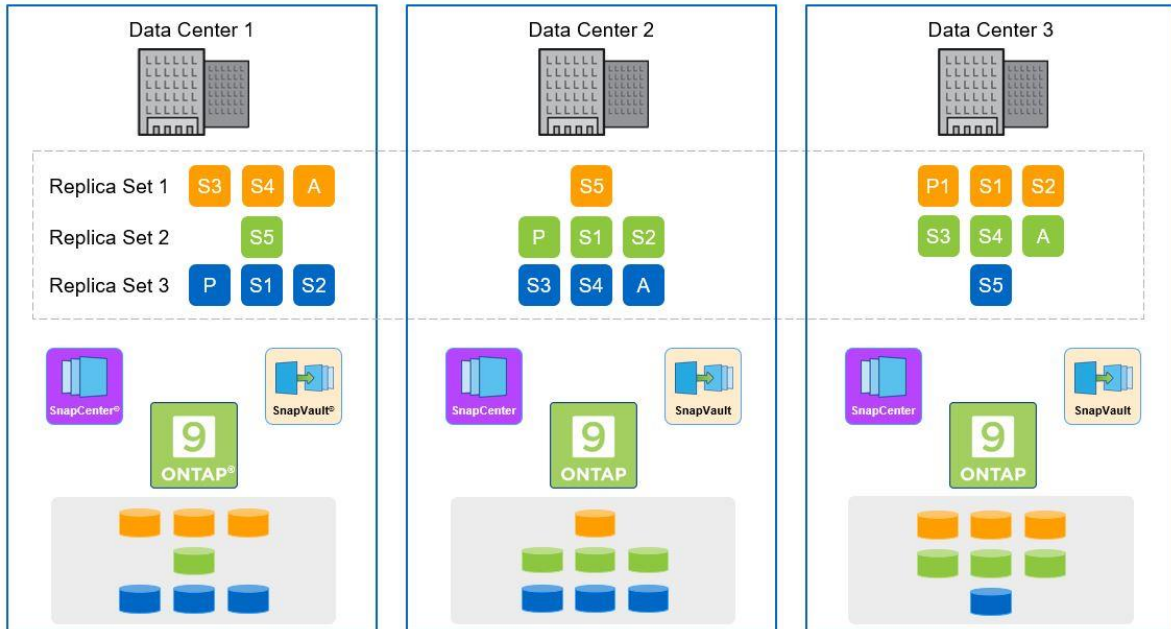
3 Customer Deployments

This section describes use cases of three customers who automated MongoDB workflows using OnCommand Workflow Automation.

3.1 Financial Services Company Saves Time and Improves Quality

A financial services company needed to reduce the time required for a MongoDB deployment, reduce the potential for error, improve the quality of service (QoS), and guarantee high availability (HA). The NetApp WFA for MongoDB solution was applied in a pilot project, resulting in a 2-week reduction in deployment time. In addition, the MongoDB QoS improved, and the automation resulted in the elimination of common deployment errors.

3.2 University Improves Disaster Recovery and Reduces Need for IT Support

A major university needed to deploy an online learning platform to increase student enrollment over a broader geographical area. The platform had to have 24/7 availability for students and span multiple geographical locations.

The university decided to implement an HA model with a 2-hour service-level agreement (SLA) for disaster recovery. By using WFA, the university was able to complete its MongoDB deployment project within 2 months.

The university network can now recover within 5 minutes of a crash. WFA also eliminated the need for 24/7 IT support. The solution is extensible and can easily be expanded to accommodate future growth.

3.3 Telecommunications Provider Rolls Out New Services Significantly Faster

To remain competitive, a telecommunications service provider needed to reduce the time required to roll out new services. The old method of manual scripting of deployment by using custom scripts was prone to error and took too long to meet customer demand. On at least one occasion, this method resulted in application bugs that could have been prevented.

The implementation of NetApp WFA for MongoDB reduced the time required for storage provisioning from 1 week to 1 hour. Other benefits included:

- Repeatable procedures that eliminated deployment errors
- Best practices automatically implemented
- Database restore time reduced from 7 days to about 5 minutes
- Greatly reduced time for deployment of dev/test environments across the company

4   MongoDB with WFA Deployment Process

The NetApp WFA for MongoDB provides a guided process and an intuitive GUI for provisioning MongoDB. WFA turns storage provisioning into a self-service, policy-based effort. This simple and effective workflow enables the MongoDB DBA to easily provision MongoDB because the storage is provisioned per the provided input. Each workflow can be repeated as needed to deploy the desired MongoDB configuration.

The following sections describe these workflow steps in detail:

1. Download the WFA template for MongoDB.
2. Add the MongoDB host.
3. Add the SVM to the data center group.
4. Provision a replica set.
5. Provision a sharded cluster.

4.1   Download the WFA Template for MongoDB

The process starts by installing the WFA software and template for MongoDB:

1. Log in to http://mysupport.netapp.com/. Select Downloads > Software from the drop-down menu.
2. Select OnCommand Workflow Automation for Linux and then click View and Download.
3. Follow the prompts to download the software and installation guide.
4. Follow the instructions to install the software and template for MongoDB.

4.2   Add the MongoDB Host

Complete these steps to add or select a MongoDB host:

1. Start the workflow to add a MongoDB host.
2. Add a new data center group or select an existing one from the drop-down menu.

   Note: The data center group is the key feature for enabling the deployment of MongoDB configurations that span multiple data centers.

3. Add the server hosts and enter the following information for each host:
   – Host Name, Host IP Address, and Host Port
   – MongoDB Server Type: Select Replicaset, Shard, or Arbiter
   – Server Protocol Type: Select iSCSI or FC
   – iSCSI Initiator Name or FC WWPN
4. Click Execute.

4.3   Add the SVM to the Data Center Group

Complete the following steps to add the Server Virtual Machine (SVM) to the data center group:

1. Start the Storage Controller Details workflow.
2. Select a data center group from the Data Center Group Name drop-down list.
3. Enter the following ONTAP SVM information:
   − SVM name
   − SVM Management IP address
4. Click Execute.

4.4 Provision a Replica Set

There are two options for provisioning MongoDB data with WFA—replica sets and sharded clusters. This section discusses replica set provisioning. Section 4.5 covers sharded clusters.

Complete the following steps to provision a replica set:

1. Start the Provision Storage workflow for the MongoDB workflow.
2. Select ReplicaSet from the MongoDB Scaling Technology drop-down list.
3. Enter the following information:
   − Size of each replica set member in gigabytes
   − Replication factor
4. Choose one server in the Primary Server(s) of Replica Set(s)/Shard(s) section.
5. Choose at least two servers in the Secondary Server(s) of Replica Set(s)/Shard(s) section.
6. Click Execute.

   Note: Arbiter, which is highlighted in yellow in the following screenshot, is always provided as a choice for the secondary replica set.

4.5 Provision a Sharded Cluster

Complete the following steps to provision a sharded cluster:

1. Start the Provision Storage workflow for the MongoDB workflow.
2. Select Sharding from the MongoDB Scaling Technology drop-down list.
3. Enter the following information:
   - Number of shards to be provisioned
   - Size of each shard in gigabytes
   - Replication factor of the replica sets to be used
   - Shard name
   - Protocol type to be used (iSCSI or FC)
   - Primary replica set members to be used by each shard
4. Click Execute.

5 Summary

MongoDB workflows are mission critical for many enterprise organizations. They give enterprises the ability to quickly extract valuable business intelligence from massive amounts of unstructured data. In today's highly competitive business environment, that intelligence is a requirement for survival and profitability. Downtime, slow performance, and data loss are completely unacceptable. Enterprise-class data storage and management are required to meet the needs of businesses using big data applications.

NetApp meets the needs of enterprise organizations running MongoDB in several ways. The NetApp AFF arrays provide the submillisecond latencies and high I/O bandwidth required by MongoDB workloads. NetApp ONTAP data management software provides a powerful set of features that enable the following:

- Near instantaneous backup and restore
- Cloud enablement for flexible scaleout and data archiving
- Storage efficiency for reduced data footprint
- Continuous availability
- Cloning for DevOps
- Data security
- Data retention for compliance

Finally, NetApp provides the automation tool for provisioning storage—OnCommand Workflow Automation (WFA) for MongoDB. As discussed and shown in the screenshots, WFA offers a repeatable, automated process for provisioning MongoDB storage using the vocabulary of MongoDB DBAs and developers. Best practices are built in, ensuring optimal storage configuration for MongoDB.

In addition, NetApp WFA for MongoDB is the industry's only multiple-data-center, end-to-end provisioning automation solution. It enables flexible deployment of MongoDB with Data Fabric for on-premises, hybrid cloud, and multicloud configurations. With NetApp ONTAP software and WFA for MongoDB, enterprise

organizations can fully utilize MongoDB and monetize vast amounts of unstructured data, enabling them to consistently win in their highly competitive business environment.

Where to Find Additional Information

To learn more about the information described in this document, refer to the following documents and websites:

- NetApp OnCommand Unified Manager
  http://www.netapp.com/us/products/management-software/oncommand/unified-manager.aspx
- NetApp OnCommand Workflow Automation 3.1 Release Notes
  https://library.netapp.com/ecm/ecm_get_file/ECMP12470051
- NetApp OnCommand Workflow Automation – Useful Workflows Cookbook
  http://www.netapp.com/us/media/ar-netapp-oncommand-workflow-automation.pdf
- NetApp TechComm TV (YouTube channel) > WFA videos
  https://www.youtube.com/results?search_query=onCommand+Workflow+Automation
- Storage Automation Store > WFA packs
  http://automationstore.netapp.com/pack-list.shtml
- TR-4492: MongoDB on the NetApp Data Fabric
  http://www.netapp.com/us/media/TR-4492.pdf
- WFA discussions
  https://community.netapp.com/t5/forums/filteredbylabelpage/board-id/oncommand-storage-management-software-discussions/label-name/workflow%20automation%20(wfa)
- WFA resources
  https://community.netapp.com/t5/OnCommand-Storage-Management-Software-Articles-and-Resources/tkb-p/oncommand-storage-management-software-articles-and-resources/label-name/workflow%20automation%20(wfa)?labels=workflow+automation+%28wfa%29
- WP-7218: NetApp Data Fabric Architecture Fundamentals: Building a Data Fabric Today
  https://cloud.netapp.com/hubfs/Data-Fabric/Data%20Fabric%20WP%20April%202017.pdf
- NetApp ONTAP 9 Documentation Center
  http://docs.netapp.com/ontap-9/index.jsp

Version History

Version	Date	Document Version History
Version 1.0	April 2018	Initial release.

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.