NetApp

eBook

Optimize Elasticsearch Performance and Costs with Cloud Volumes ONTAP

Elasticsearch: A New Approach to Search and Analytics

With the growth in the volume and increasing complexity of production deployments and operational pipelines, managing logs and metrics is a crucial component of successful IT deployment. DevOps teams and administrators need logs and metrics to gain operational insights, meet their SLA obligations, prevent unauthorized access, and to identify errors, anomalies, or suspicious activity. For that, many companies turn to Elasticsearch.

Elasticsearch is a distributed search engine and data analytics platform based on the Apache Lucene search and indexing library that is orders of magnitude faster than traditional SQL databases. This speed is due to the advanced inverted index data structures Elasticsearch uses to store data. It also provides advanced search functionality (autosuggestion, percolators, “did-you-mean” functionality), language analyzers and parsers, data aggregation, and statistical metrics. On top of that, Elasticsearch allows data to scale across cluster nodes with sharding and provides advanced replication, cluster management, and index management features.

These features make Elasticsearch powerful for analyzing different types of structured and unstructured data, including full text, geospatial, and numerical data. Elasticsearch is also widely used for storing logs and metrics due to its advanced data aggregation features, easy-to-use REST API and query language, fast granular search, and smooth integration with popular logging and metrics shippers such as Logstash and Metricbeat. Logs and metrics can be visualized in Kibana, — the data analysis and visualization tool also developed by Elastic and being part of a popular ELK (Elasticsearch, Logstash, Kibana) stack.

The Challenge

Elasticsearch is one of the most efficient systems for collecting, indexing, and unifying log, metrics, and other mission-critical data across different environments. However, as with any software system, many challenges arise when running Elasticsearch in production at scale. Some of these challenges are:

Performance

With the growing volumes of data and its distributed storage (e.g., shards in Elasticsearch), ensuring low search latency becomes more and more challenging. Efficient and fast search requires agile storage infrastructure, network configuration, and optimization of every bit of your system.

Compute and Storage Costs

As volumes of data are ingested into your production clusters, more high-performance storage and compute is required, increasing costs.

Storage over-provisioning

With most cloud providers, companies need to pre-provision vast amounts of block storage which may stay idle during low traffic times. Under-utilization of storage entails unnecessary costs that can be avoided.

Data protection

Data needs to be protected from application and network failures and downtimes with efficient snapshots, backup, and disaster recovery (DR) mechanisms.

Data retention

Due to high storage costs, companies adopt shorter log retention time windows which limit the flexibility of their data management strategy.

The Solution

How can these problems be solved? NetApp Cloud Volumes ONTAP addresses the performance and cost challenges that come with high-volume Elasticsearch deployments leading to dramatic improvements in performance and considerable reductions of storage costs and operational overheads for your Elasticsearch clusters.

Key Benefits

Reduce operational overhead

- No need to re-index data when the cluster scales
- Create Cross-Region Replication in seconds
- Clone clusters in seconds

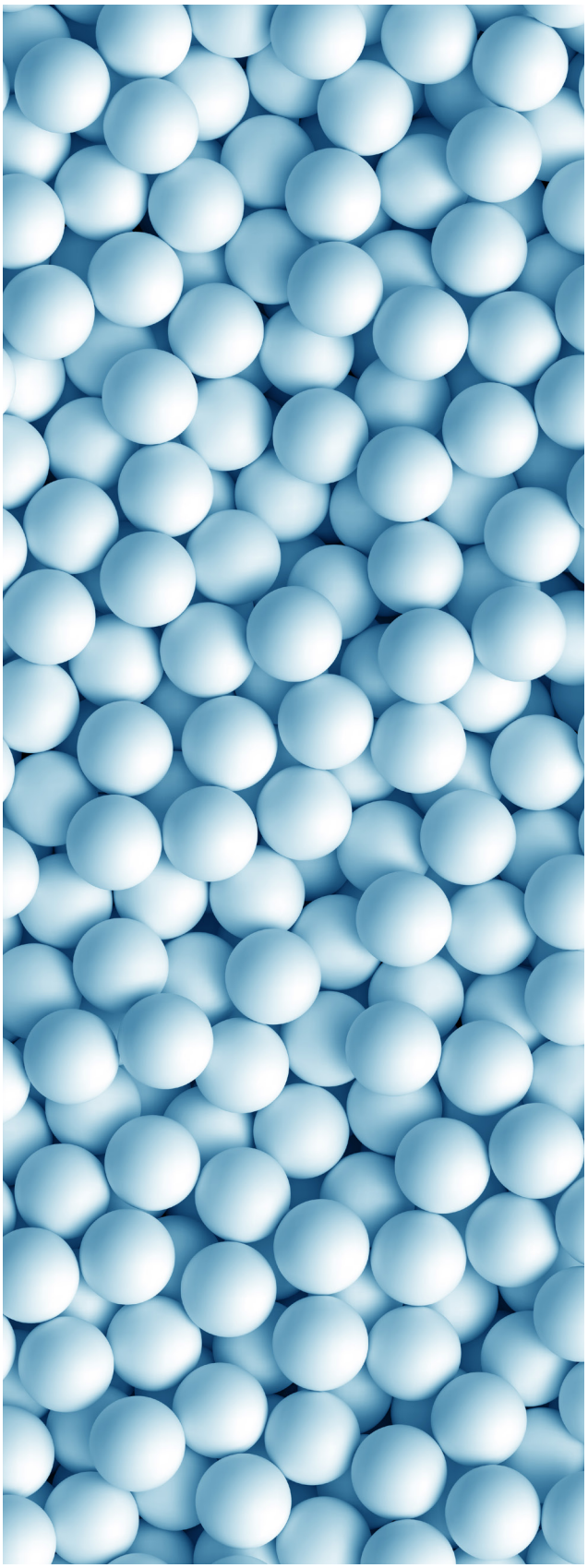
Reduce costs

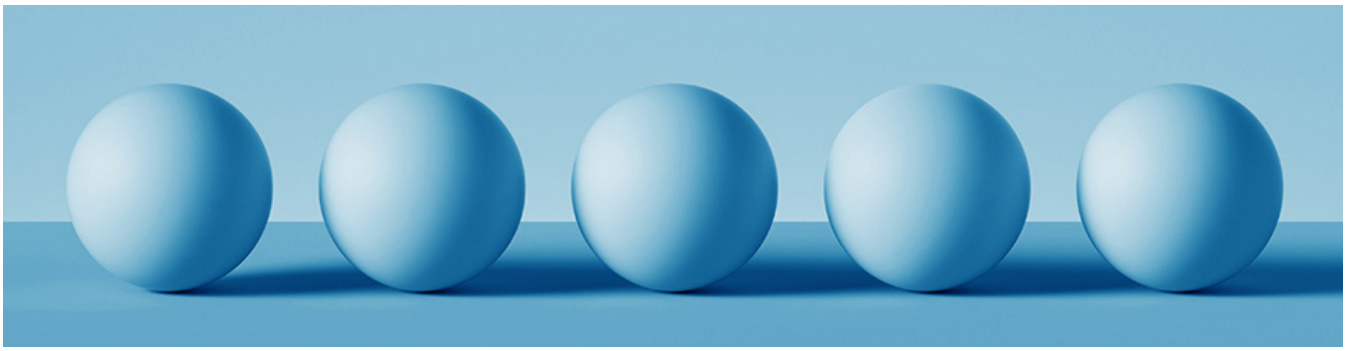
- Optimize write-intensive workloads costs
- **40%** storage cost reduction
- Tier infrequently-used cold data to less-expensive object storage
- Clone clusters with zero cost
- Sustain high performance while reducing spend

Increase performance of read-intensive workloads*

- **66%** faster data querying
- **86%** more IOPS
- **223%** more throughput

*See the benchmark testing details in the Performance section below. →





NetApp Cloud Volumes ONTAP

NetApp's Cloud Volumes ONTAP software-defined storage management solution supports up to a capacity of 2 PBs, and supports various use cases such as file services, databases, DevOps or any other enterprise workload, on AWS, Azure, and Google Cloud.

Cloud Volumes ONTAP provides many enterprise-grade storage management features on top of cloud storage:

1 High availability

The two-node high availability configuration protects against cluster and network outages keeping RPO at zero and RTO below 60 seconds.

2 Data Mobility

SnapMirror® data replication seamlessly replicates data for disaster recovery, testing, and cloud migration.

3 Data protection

NetApp Snapshot™ copies are cost and space-efficient snapshots of your data that can be restored seamlessly, protecting against data loss and/or corruption.

4 Hybrid and Multicloud

Manage and maintain data seamlessly across different clouds and on-prem data centers with the easy-to-use NetApp Cloud Manager interface and API.

5 Automation

RESTful API calls let developers treat Infrastructure as Code (IAC), significantly speeding up the dev/test pipeline and decreasing time to market (TTM).

6 Storage efficiencies

Cloud Volumes ONTAP storage efficiency features allow companies to cut up to 70% of their cloud storage costs without affecting performance.

7 Data tiering

Identifies infrequently-used data and moves it automatically between high performance block storage and less-expensive object storage.

8 Multiprotocol Access

Support for different storage protocols including NFS, SMB and iSCSI, allowing the same storage to be configured for file share across the network as well as block storage.

9 High Performance

WAFL and cloud data caching secure enhanced storage throughput for performance-intensive workloads.

10 Kubernetes and Container Integration

Cloud Volumes ONTAP can be used with NetApp Trident to provision persistent storage for containerized stateful applications.

11 Security and Safety

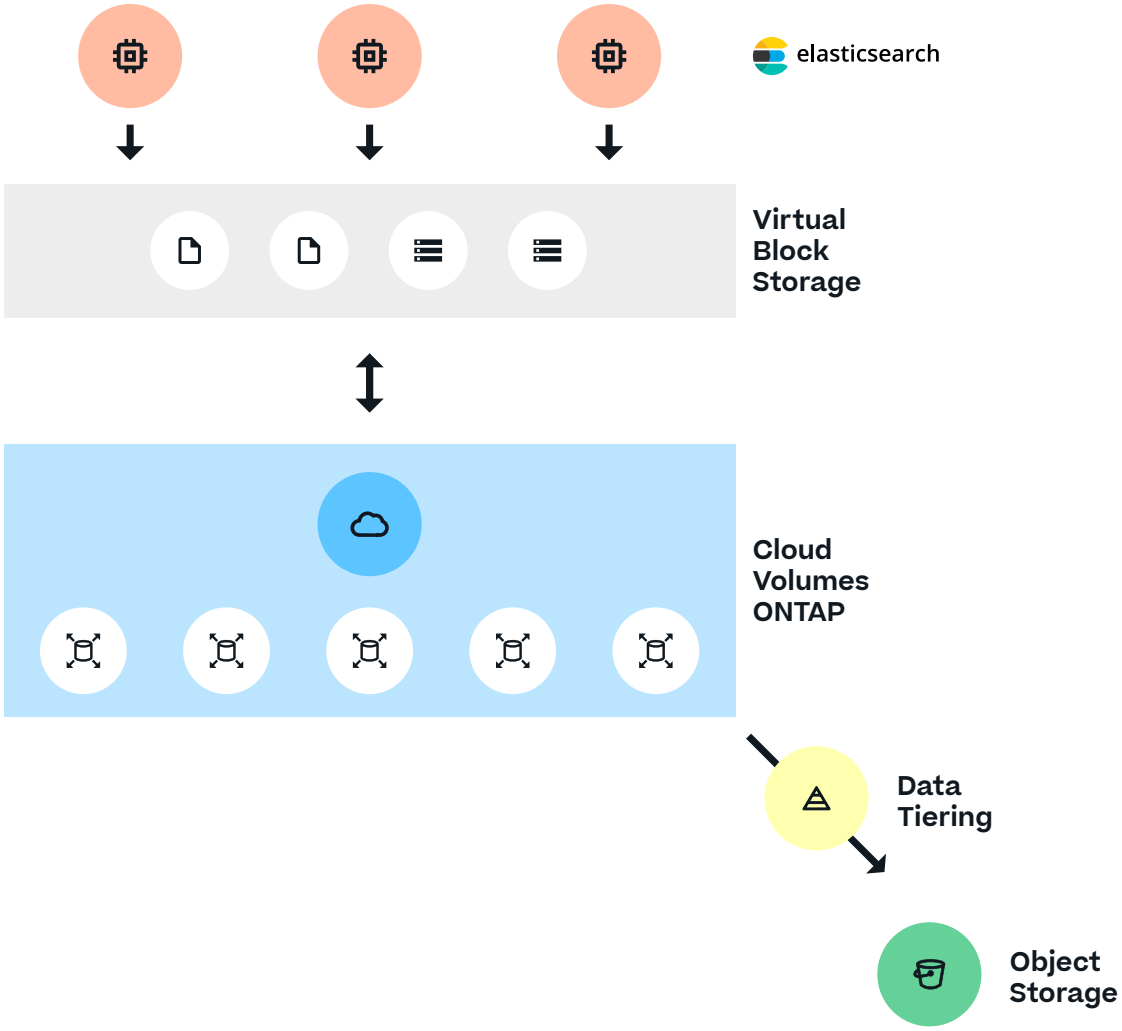
Cloud Volumes ONTAP offers storage security features such as ransomware protection, encryption at-rest and in-transit, VNET integration for perimeter security, and cloud WORM (write once, read many) storage.

12 Compliance

The NetApp Cloud Compliance add-on can scan cloud data to identify sensitive private data that could fall under regulatory scope of GDPR, CCPA, and other data privacy legislation.

Even though Elasticsearch is highly optimized for fast search performance, that performance depends on the underlying storage used by Elasticsearch. When using Cloud Volumes ONTAP together with Elasticsearch, Cloud Volumes ONTAP serves as a storage management layer for the cloud-based storage volumes (e.g., Amazon EBS, Google Persistent Disk, or Azure Disks) used by Elasticsearch to store and manage data.

Elasticsearch with Cloud Volumes ONTAP Architecture Diagram



Elasticsearch with Cloud Volumes ONTAP

Performance	Cost-Optimization	Operational Overhead Reduction

Performance

High search performance is a necessary prerequisite for fast user access to important data as well as DevOps access to business-critical information such as logs and metrics. Being able to search quickly is also required for implementing advanced search features of Elasticsearch, such as “did-you-mean” functionality, auto-suggestion, and fuzzy queries, all of which require near-instantaneous access to indexes.

High search latency can result in a deteriorated customer experience and inability to meet SLA obligations to clients.

Cloud Volumes ONTAP improves Elasticsearch search performance. In particular, Cloud Volumes ONTAP dramatically decreases query time for accessing Elasticsearch indexes.

To demonstrate these benefits, we conducted an in-house benchmark test of Cloud Volumes ONTAP iSCSI gp2 versus Amazon EBS gp2 volumes on logs and metrics data that simulated Kibana searches against Elasticsearch indexes and data aggregation (metrics, buckets, pipelines, etc.) used to create interactive Kibana dashboards. We ran this test against the 420 GB Elasticsearch index composed of 14 30GB daily indexes of logs obtained from various sources.

Watch this presentation about the power of Elasticsearch running on Cloud Volumes ONTAP Storage:

The benchmark test found that Cloud Volumes ONTAP-managed volumes can achieve 86 percent more IOPS and 223 percent more throughput compared to standard EBS gp2 volumes. This is a significant improvement in IOPS and throughput which are especially important for the speed of data aggregation operations with multiple fields and write-intensive operations on logs and metrics. This improved IOPS can dramatically enhance search performance for Elasticsearch clusters compared to the native AWS block storage.

Cost-Optimization

Cloud storage pricing at scale can be very high at times, especially for highly performant storage. This can put a strain on the success of your company cloud-based operations. The burden of storage costs is especially noticeable for write-intensive Elasticsearch clusters that consume large volumes of logs, metrics, and other data.

Cloud Volumes ONTAP's storage efficiency features reduce the storage footprint and costs of companies using cloud storage. Some of the Cloud Volumes ONTAP features that lead to dramatic storage costs reduction are the following:

Thin Provisioning
Just-in-time, on-demand provisioning of disk space so you don't need to pay for unused storage.

Data Tiering
Automatically tiers data between less-expensive object storage and the high-performance block storage. In Elasticsearch, cold indexes can be tiered while hot indexes (e.g. recent logs) remain on the more performant storage.

Deduplication
Frees up space occupied by duplicates of the same data blocks, so you don't pay twice for the same data.

Compression
Automatically compresses data blocks, reducing the amount of storage space consumed, saving up to 70%.

Compaction
Automatically combines remaining small chunks of data into single 4 KB physical blocks to improve storage efficiency.

NetApp Snapshot™ Copies
A more storage-efficient and flexible alternative to standard, native cloud snapshots.
In addition, SnapVault® and SnapMirror® tools enable data replication and backup to or from the cloud.

Data Cloning
FlexClone® technology creates instant, zero-cost penalty, writable clones of cluster data that only consume space for delta data.

Operational Overhead Reduction

Operational overheads in Elasticsearch deployments are typically failures to automate important processes and put in place policies and procedures that prevent disruptions. For example, manually restoring Elasticsearch data from a snapshot is a time-consuming process that requires DevOps and Elasticsearch administrator involvement, which could be avoided with sound snapshot and backup policies. Operational overheads do not only squeeze valuable time off your team but lead to costly service disruptions and downtimes. Cloud Volumes ONTAP will help you avoid such problems.

Enabling High Availability

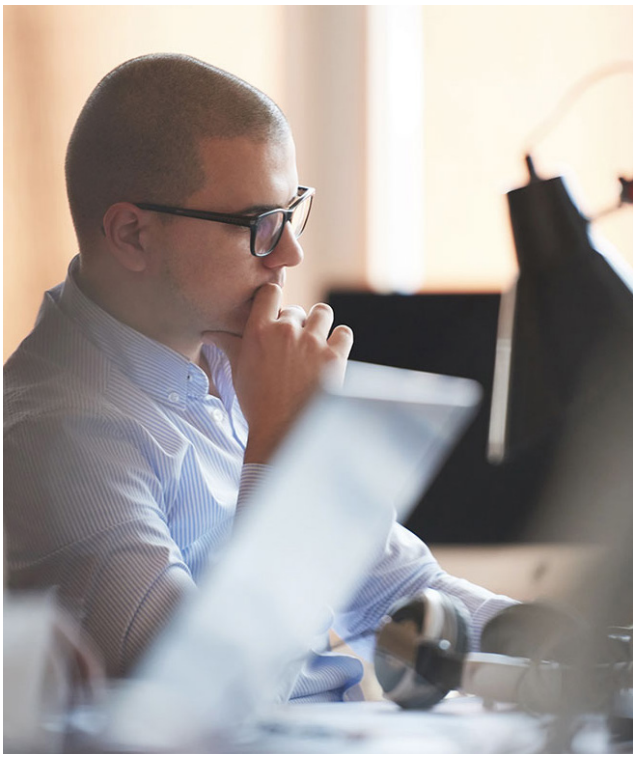
Elasticsearch has a built-in replication feature for high availability (HA). However, to ensure full data protection we need yet another layer of HA for the underlying cloud storage.

Cloud Volumes ONTAP's high availability configuration creates an additional layer of data protection by generating two storage environments and writing Elasticsearch data simultaneously to both, ensuring zero RPO and RTO below 60 seconds. Data is synchronously mirrored between the two nodes in the deployment so if one node fails, the other instantly takes over. Once the failed node reboots, the second node will resync data and the system returns to dual-node operation.

Cluster Cloning with NetApp FlexClone®

Setting up dev/test environments for Elasticsearch clusters can lead to significant operational overheads.

Cloud Volumes ONTAP uses FlexClone to create zero-capacity, writable clones of any storage volume whether on-prem or in the cloud. The clones of your data are created instantly, no matter the size of the underlying volumes because a FlexClone volume and its parent volume share the same disk space for common data—additional space is only required when changes are made to volume clones. This makes FlexClone a powerful tool for the fast deployment of testing environments, speeding CI/CD processes and TTM while also cutting storage costs.



Improved Data Protection

Snapshots, backup, and disaster recovery are crucial components of data protection for Elasticsearch. But Elasticsearch's Snapshot and Restore API works only at the application level, which doesn't extend to the underlying storage layer (e.g. block storage).

NetApp Snapshot copies are application-consistent due to their direct integration with the underlying applications (e.g. database) snapshot and restore software and use built-in space-efficiencies to reduce storage consumption.

To automate the disaster recovery process, NetApp offers an efficient disaster recovery (DR) mechanism implemented in the SnapMirror tool. SnapMirror creates a standby DR environment fully synchronized with a production Elasticsearch cluster so if a cluster goes offline, SnapMirror can automatically fail over to a standby that runs with a warm or hot copy of your Elasticsearch data. SnapMirror incrementally updates the standby cluster synchronously, which is more efficient than full point-in-time snapshots and consumes less network bandwidth and storage thanks to Cloud Volumes ONTAP's space efficiencies.

A More Affordable and Better Protected Way to Use Elasticsearch

Cloud Volumes ONTAP is a powerful cloud storage management tool that provides a lot of enterprise-grade features for cloud storage deployments. Its built-in performance and storage efficiency features dramatically improve the Elasticsearch deployments in the cloud. High availability, data protection, and data cloning features also decrease operational overheads for managing storage in the cloud, making it easy for DevOps to ensure the continuing effective operation of Elasticsearch deployments.

If you are running Elasticsearch in the cloud and want to decrease storage costs, sign up for a free trial of Cloud Volumes ONTAP now.

Start a free trial of Cloud Volumes ONTAP now

Start now

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.

NA-000-1120