



Performance overview and validation with AFF on-premises

NetApp Solutions

NetApp
October 20, 2023

This PDF was generated from <https://docs.netapp.com/us-en/netapp-solutions/data-analytics/kafka-nfs-performance-overview-and-validation-with-aff-on-premises.html> on October 20, 2023. Always check docs.netapp.com for the latest.

Table of Contents

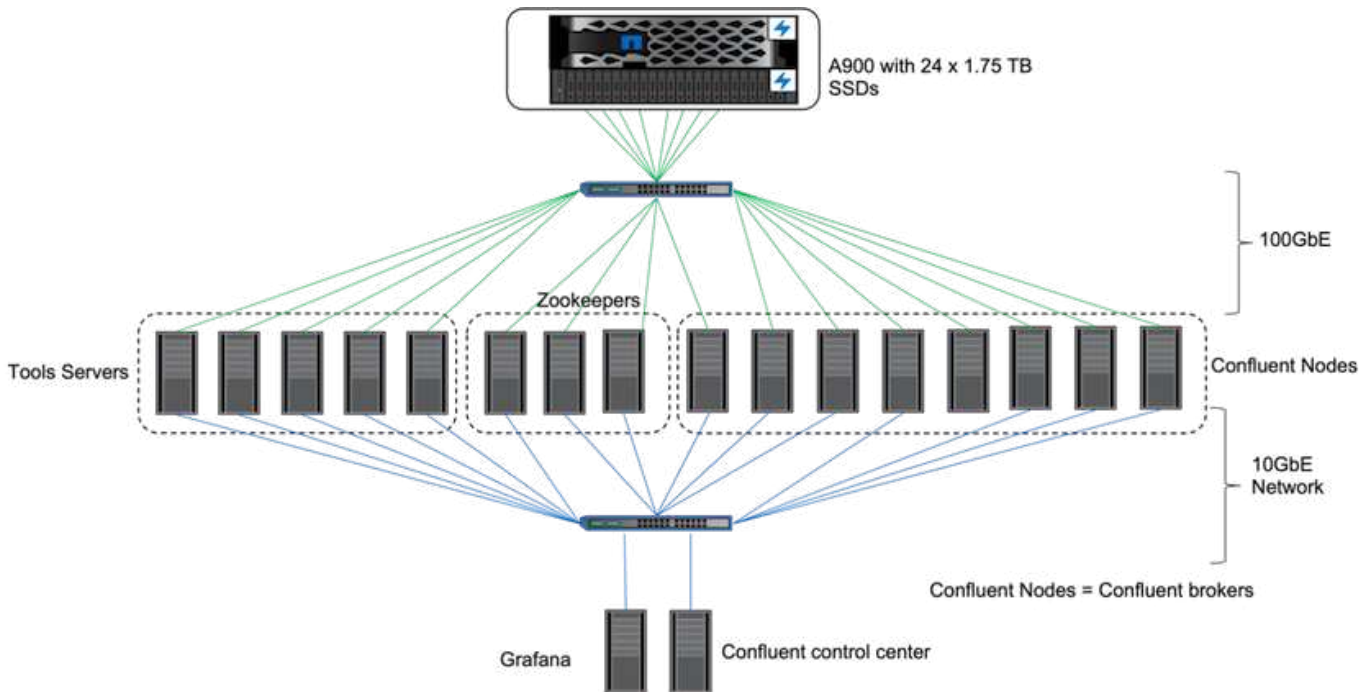
- Performance overview and validation with AFF A900 on-premises 1
 - Storage configuration 1
 - Client tuning 1
 - Kafka broker tuning 2
 - Workload generator testing methodology 2
 - Extreme performance and exploring storage limits 5
 - Sizing guidance 6

Performance overview and validation with AFF A900 on-premises

[Previous: Performance overview and validation in AWS.](#)

On-premises, we used the NetApp AFF A900 storage controller with ONTAP 9.12.1RC1 to validate the performance and scaling of a Kafka cluster. We used the same testbed as in our previous tiered storage best practices with ONTAP and AFF.

We used Confluent Kafka 6.2.0 to evaluate the AFF A900. The cluster features eight broker nodes and three zookeeper nodes. For performance testing, we used five OMB worker nodes.



Storage configuration

We used NetApp FlexGroups instances to provide a single namespace for log directories, simplifying recovery and configuration. We used NFSv4.1 and pNFS to provide direct path access to log segment data.

Client tuning

Each client mounted the FlexGroup instance with the following command.

```
mount -t nfs -o vers=4.1,nconnect=16 172.30.0.121:/kafka_vol01  
/data/kafka_vol01
```

In addition, we increased the `max_session_slots`` from the default 64 to 180. This matches the default session slot limit in ONTAP.

Kafka broker tuning

To maximize throughput in the system under test, we significantly increased the default parameters for certain key thread pools. We recommend following Confluent Kafka best practices for most configurations. This tuning was used to maximize the concurrency of outstanding I/O to storage. These parameters can be adjusted to match your broker's compute resources and storage attributes.

```
num.io.threads=96
num.network.threads=96
background.threads=20
num.replica.alter.log.dirs.threads=40
num.replica.fetchers=20
queued.max.requests=2000
```

Workload generator testing methodology

We used the same OMB configurations as for cloud testing for the Throughput driver and topic configuration.

1. A FlexGroup instance was provisioned using Ansible on an AFF cluster.

```

---
- name: Set up kafka broker processes
  hosts: localhost
  vars:
    ntap_hostname: 'hostname'
    ntap_username: 'user'
    ntap_password: 'password'
    size: 10
    size_unit: tb
    vservers: vs1
    state: present
    https: true
    export_policy: default
    volumes:
      - name: kafka_fg_vol01
        aggr: ["aggr1_a", "aggr2_a", "aggr1_b", "aggr2_b"]
        path: /kafka_fg_vol01
  tasks:
    - name: Edit volumes
      netapp.ontap.na_ontap_volume:
        state: "{{ state }}"
        name: "{{ item.name }}"
        aggr_list: "{{ item.aggr }}"
        aggr_list_multiplier: 8
        size: "{{ size }}"
        size_unit: "{{ size_unit }}"
        vservers: "{{ vservers }}"
        snapshot_policy: none
        export_policy: default
        junction_path: "{{ item.path }}"
        qos_policy_group: none
        wait_for_completion: True
        hostname: "{{ ntap_hostname }}"
        username: "{{ ntap_username }}"
        password: "{{ ntap_password }}"
        https: "{{ https }}"
        validate_certs: false
        connection: local
        with_items: "{{ volumes }}"

```

2. pNFS was enabled on the ONTAP SVM.

```
vserver modify -vservers vs1 -v4.1-pnfs enabled -tcp-max-xfer-size 262144
```

3. The workload was triggered with the Throughput driver using with same workload configuration as for Cloud Volumes ONTAP. See the section [“Steady state performance”](#) below. The workload used a replication factor of 3, meaning three copies of log segments were maintained in NFS.

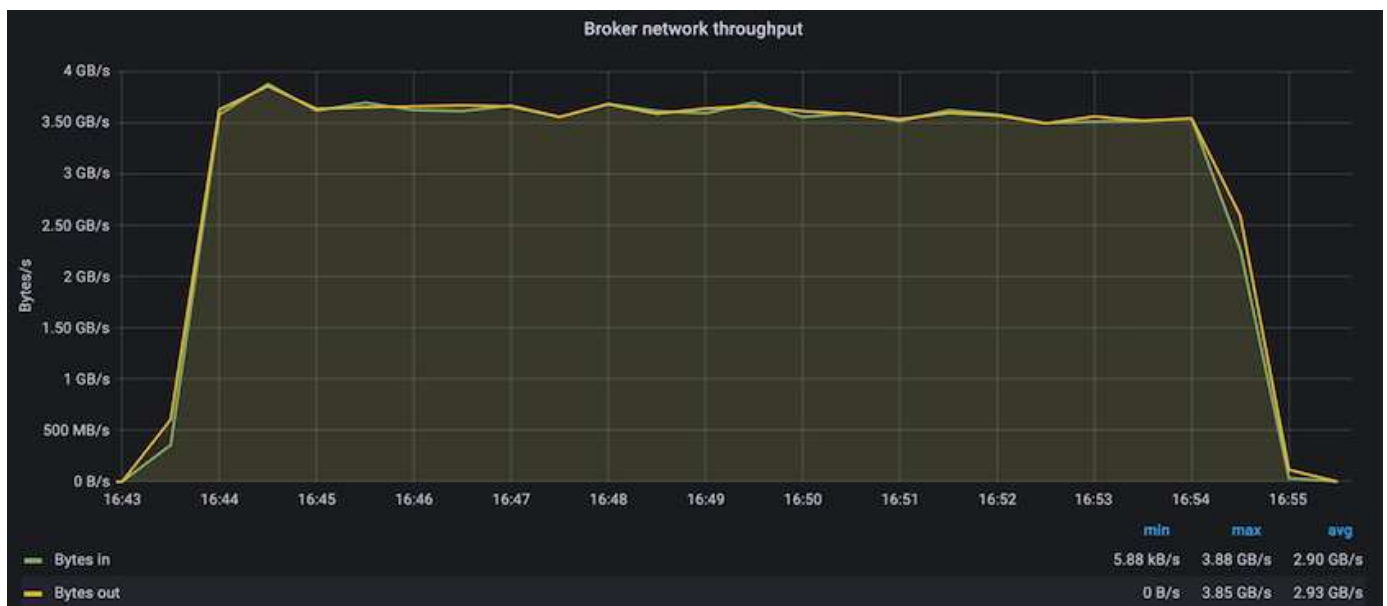
```
sudo bin/benchmark --drivers driver-kafka/kafka-throughput.yaml  
workloads/1-topic-100-partitions-1kb.yaml
```

4. Finally, we completed measurements using a backlog to measure the ability of consumers to catch up to the latest messages. OMB constructs a backlog by pausing consumers during the beginning of a measurement. This produces three distinct phases: backlog creation (producer-only traffic), backlog draining (a consumer-heavy phase in which consumers catch up on missed events in a topic), and the steady state. See the section [“exploring storage limits”](#) for more information.

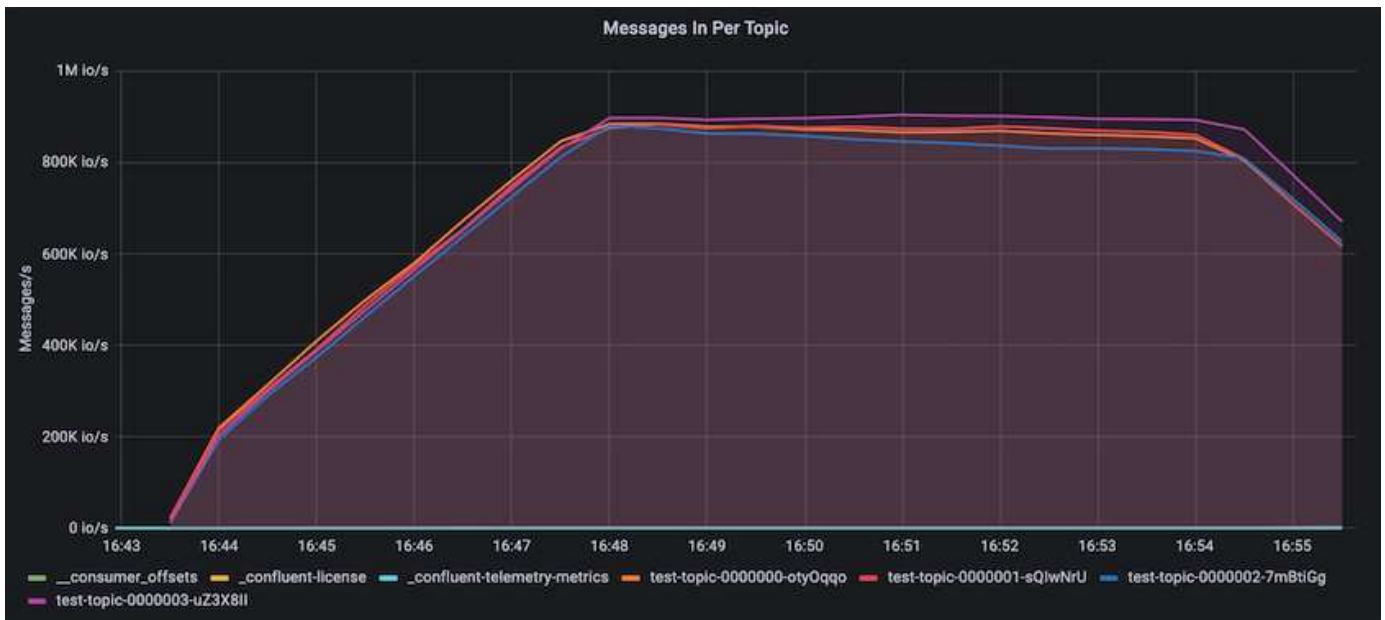
Steady state performance

We evaluated the AFF A900 using the OpenMessaging Benchmark to provide a similar comparison as for Cloud Volumes ONTAP in AWS and DAS in AWS. All performance values represent Kafka-cluster throughput at the producer and consumer level.

Steady state performance with Confluent Kafka and the AFF A900 achieved over 3.4GBps average throughput for both producer and consumers. This is over 3.4 million messages across the Kafka cluster. By visualizing the sustained throughput in bytes per second for BrokerTopicMetrics, we see the excellent steady state performance and traffic supported by the AFF A900.



This aligns well with the view of messages delivered per topic. The following graph provides a per-topic breakdown. In the configuration tested we saw nearly 900k messages per topic across four topics.

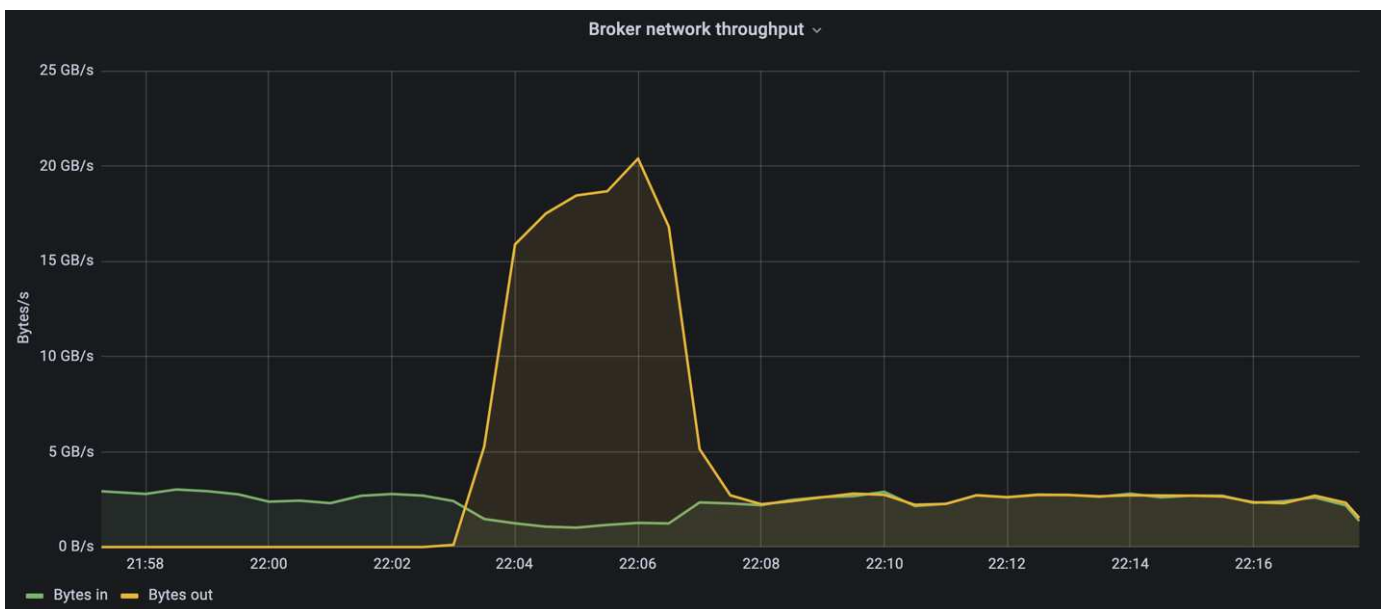


Extreme performance and exploring storage limits

For AFF, we also tested with OMB using the backlog feature. The backlog feature pauses consumer subscriptions while a backlog of events is built up in the Kafka cluster. During this phase, only producer traffic occurs, which generates events that are committed to logs. This most closely emulates batch processing or offline analytics workflows; in these workflows, consumer subscriptions are started and must read historical data that has already been evicted from the broker cache.

To understand the storage limitations on consumer throughput in this configuration, we measured the producer-only phase to understand how much write traffic the A900 could absorb. See the next section [“Sizing guidance”](#) to understand how to leverage this data.

During the producer-only part of this measurement, we saw high peak throughput that pushed the limits of A900 performance (when other broker resources were not saturated serving producer and consumer traffic).





We increased the message size to 16k for this measurement to limit per-message overheads and maximize storage throughput to NFS mount points.

```
messageSize: 16384
consumerBacklogSizeGB: 4096
```

The Confluent Kafka cluster achieved a peak producer throughput of 4.03GBps.

```
18:12:23.833 [main] INFO WorkloadGenerator - Pub rate 257759.2 msg/s /
4027.5 MB/s | Pub err      0.0 err/s ...
```

After OMB completed populating the eventbacklog, consumer traffic was restarted. During measurements with backlog draining, we observed peak consumer throughput of over 20GBps across all topics. The combined throughput to the NFS volume storing the OMB log data approached ~30GBps.

Sizing guidance

Amazon Web Services offers a [sizing guide](#) for Kafka cluster sizing and scaling.

This sizing provides a useful formula for determining storage throughput requirements for your Kafka cluster:

For an aggregated throughput produced into the cluster of $t_{cluster}$ with a replication factor of r , the throughput received by the broker storage is as follows:

$$t_{storage} = t_{cluster} / \#brokers + t_{cluster} / \#brokers * (r-1)$$
$$= t_{cluster} / \#brokers * r$$

This can be simplified even further:

$$\max(t_{cluster}) \leq \max(t_{storage}) * \#brokers / r$$

Using this formula allows you to select the appropriate ONTAP platform for your Kafka hot tier needs.

The following table explains the anticipated producer throughput for the A900 with different replication factors:

Replication factor	Producer throughput (GPps)
3 (measured)	3.4
2	5.1
1	10.2

[Next: Conclusion.](#)

Copyright information

Copyright © 2023 NetApp, Inc. All Rights Reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP “AS IS” AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

LIMITED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (b)(3) of the Rights in Technical Data -Noncommercial Items at DFARS 252.227-7013 (FEB 2014) and FAR 52.227-19 (DEC 2007).

Data contained herein pertains to a commercial product and/or commercial service (as defined in FAR 2.101) and is proprietary to NetApp, Inc. All NetApp technical data and computer software provided under this Agreement is commercial in nature and developed solely at private expense. The U.S. Government has a non-exclusive, non-transferrable, nonsublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b) (FEB 2014).

Trademark information

NETAPP, the NETAPP logo, and the marks listed at <http://www.netapp.com/TM> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.