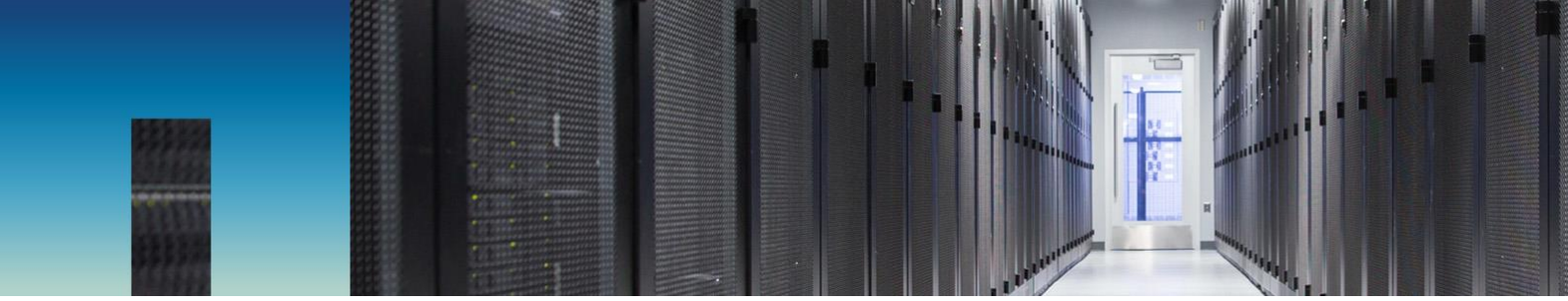


Technical Report

NetApp ONTAP AI and OmniSci GPU-Accelerated Analytics Platform

ONTAP in an OmniSci Environment

Karthikeyan Nagalingam, NetApp
Jason Bachman, OmniSci

August 2019 | TR-4793

Abstract

This document covers a deployment solution for OmniSci and NetApp®, including a NetApp storage architecture, NetApp storage efficiency features such as deduplication and compression, and backup with NetApp Snapshot™ technology. It also discusses the use of NetApp FlexClone® technology to create OmniSci graphics processing unit (GPU) databases for development and testing use cases.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

This document covers a deployment solution for OmniSci and NetApp, including a NetApp storage architecture, NetApp storage efficiency features such as deduplication and compression, and backup with NetApp Snapshot technology. It also discusses the use of NetApp FlexClone technology to create OmniSci graphics processing unit (GPU) databases for development and testing use cases.

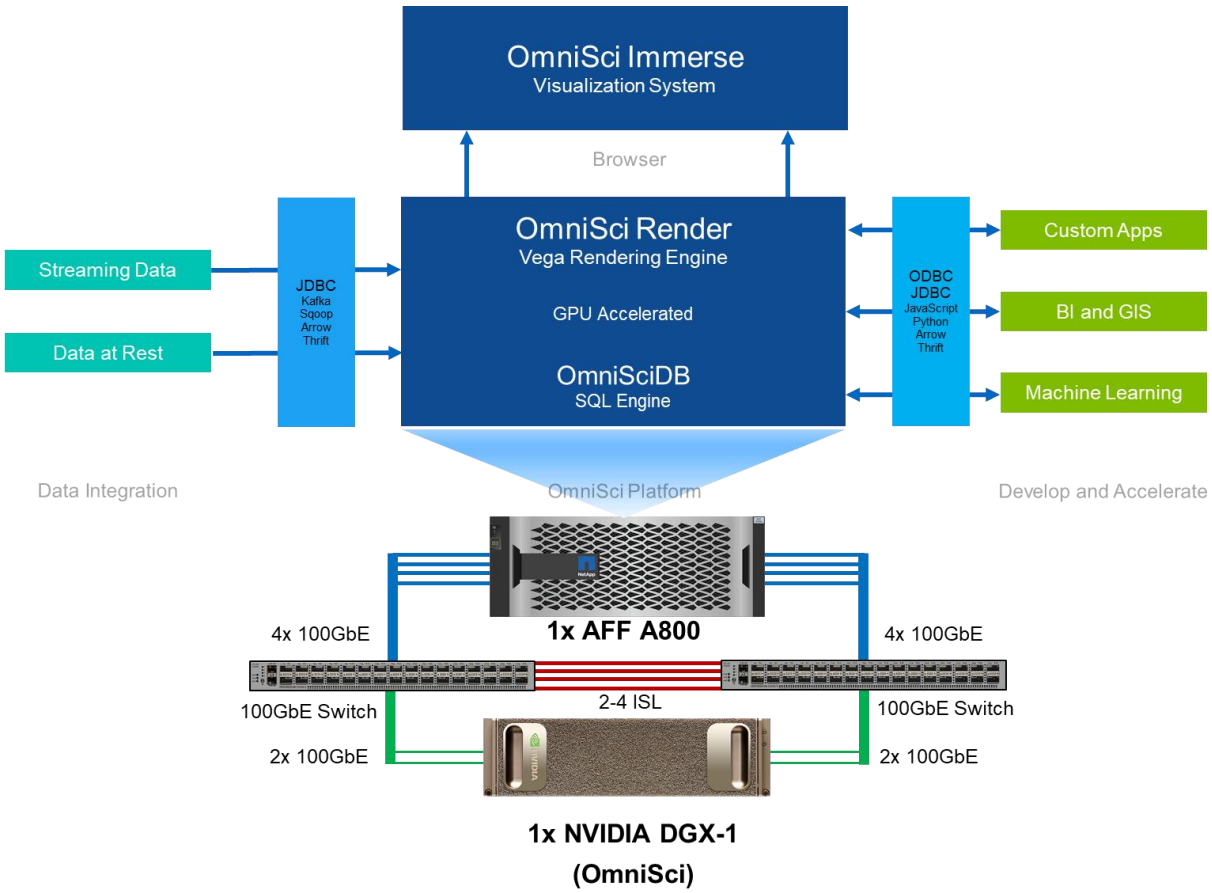
1.1 Target Audience

This document is intended for data scientists, geospatial analysts, and other users of big data analytics. This solution will help you interact with large volumes of data in a shorter time frame and will help you with the design and management of machine learning models. This solution will enable you to interactively explore up to millions of polygons and billions of mapped points. This solution will also help administrators and developers install and configure an OmniSci instance, load data for analysis, and extend OmniSci custom charts and interfaces.

1.2 Solution Architecture

The NetApp ONTAP® AI and OmniSci solution is designed to handle large datasets by using the processing power of GPUs alongside traditional CPUs to achieve high performance and optimal data protection and management. Figure 1 presents an overview of the architecture.

Figure 1) Architecture overview.



This NetApp ONTAP AI and OmniSci solution provides the following key benefits:

- High performance
- Storage efficiency to meet cost objectives
- Robust data protection to meet low recovery point objectives and low recovery time objectives with no data loss

- Full scalability of data storage

2 Test Plan Summary

To test this storage solution, we ran the OmniSci platform with 9 billion records and visualized geospatial data points. NetApp storage technologies such as Snapshot copies, cloning, and storage efficiency processes were deployed on top of the OmniSci dataset.

3 Test Results Summary

Table 1 summarizes the results from all the tests that we performed in support of this solution.

Table 1) Test result summary.

Test Description	Result Summary
Load geospatial data.	Data was successfully ingested into an OmniSciDB database that stores data with ONTAP AI.
Create dashboards and visualize data.	Creation and visualization were successful.
Create NetApp Snapshot copies.	No visible effects on performance were observed.
Create NetApp FlexClone volumes.	There was no observed effect on OmniSci performance. Clone configuration for the OmniSciDB database was completed.

Note: For the Snapshot copy creation test, we created a consistency group (CG) Snapshot copy.

4 Technology Overview

4.1 NetApp ONTAP AI Powered by ONTAP 9

NetApp ONTAP AI, a proven architecture powered by NVIDIA DGX supercomputers and NetApp cloud-connected storage, has been developed and verified by NetApp and NVIDIA. It provides your organization with a prescriptive architecture that delivers the following benefits:

- It eliminates design complexities.
- It permits the independent scaling of compute and storage.
- It can start small and scale seamlessly.
- It provides a range of storage options for various performance and cost points.

ONTAP AI integrates NVIDIA DGX-1 servers with NVIDIA Tesla V100 GPUs and a NetApp AFF A800 system with state-of-the-art networking. ONTAP AI simplifies artificial intelligence (AI) deployments by eliminating design complexity and guesswork. Your enterprise can start small and grow nondisruptively while intelligently managing data from the edge to the core to the cloud and back.

NetApp ONTAP 9 data management software is an optimal solution for a GPU database such as OmniSciDB. ONTAP is the enterprise data management software that powers NetApp AFF and FAS systems and the software-only NetApp Cloud Volumes ONTAP.

Storage Efficiency

ONTAP 9 offers inline deduplication, compression, and compaction. Whether it's written to on-premises or cloud storage, data occupies less space, which translates to lower data storage costs.

Snapshot Copies and FlexClone

With ONTAP Snapshot technology, you can create point-in-time data copies with no effect on performance and with minimal consumption of storage space. You can create these Snapshot copies almost instantaneously and use them with NetApp SnapRestore® software to recover entire file systems or data volumes in seconds.

4.2 OmniSci Overview

OmniSci (formerly MapD) is the pioneer in GPU-accelerated analytics. The OmniSci platform is used in business and in government to find insights in data beyond the limits of mainstream analytics tools. Harnessing the massive parallel computing power of GPUs, the platform is available in the cloud (OmniSci Cloud and the leading public cloud providers) and on the premises. OmniSci is available in Enterprise Edition and OmniSciDB Open Source Edition. OmniSci originated from research at Harvard and the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL). OmniSci is funded by GV, In-Q-Tel, New Enterprise Associates (NEA), NVIDIA, Tiger Global Management, Vanedge Capital, and Verizon Ventures. The company is headquartered in San Francisco. Learn more about OmniSci at www.OmniSci.com.

OmniSci GPU Database

OmniSciDB is the foundation of the OmniSci platform. OmniSciDB is SQL-based, relational, and columnar, and it was specifically developed to harness the parallel processing power of GPUs. OmniSciDB can query up to billions of rows in milliseconds, and is capable of unprecedented ingestion speeds, making it the ideal SQL engine for the era of big, high-velocity data. Find more information about OmniSciDB here.

5 Test Configuration Details

This section describes the tested configurations, the network infrastructure, OmniSci running DGX-1, and the storage provisioning details.

5.1 Solution Architecture

For the validation, we used the solution components that are listed in Table 2.

Table 2) Base components for the solution architecture.

Solution Components	Details
OmniSci Server: 4.6.1-20190429-02ec2e206b	• One NVIDIA DGX-1 server
Linux (Ubuntu 18.04.2 LTS)	
NetApp AFF array high-availability (HA) pair	• NetApp ONTAP 9 software • 24 x 960GB SSDs • NFS protocol • NetApp AFF A800 array

In this validation, we used one DGX-1 for OmniSci software installation. The OmniSci software, data, and logs were stored in NetApp AFF A800 storage with the NFS protocol. Figure 2 shows the tested configuration topology, and Table 3 provides details.

Figure 2) Network topology of tested configuration.

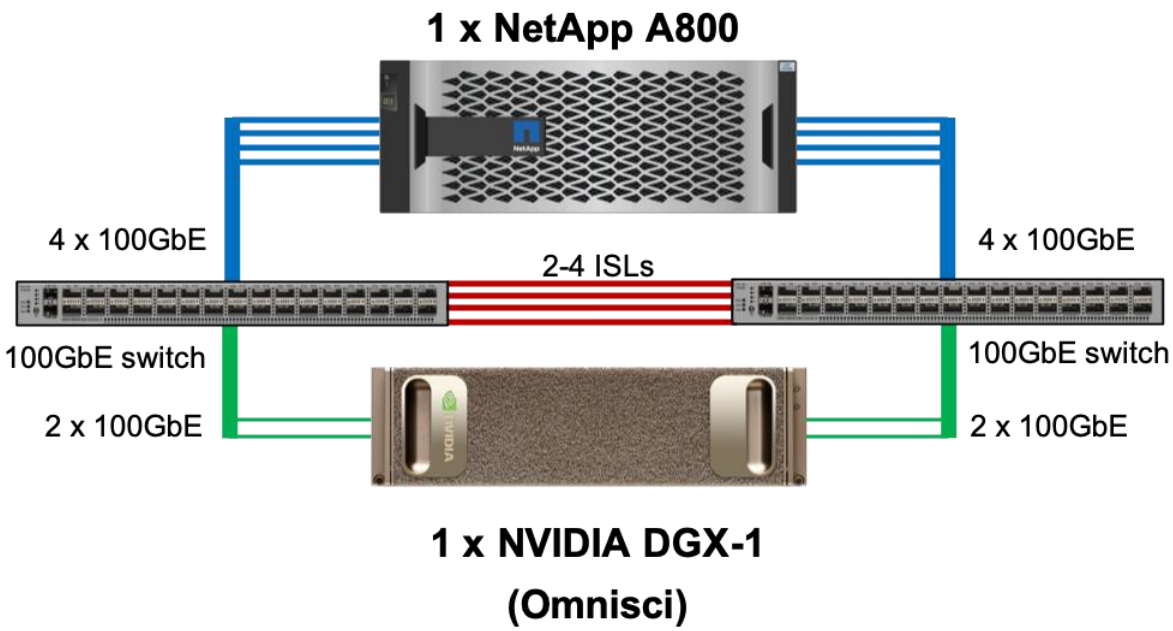


Table 3) Storage configuration.

Controller	Aggregate	Volume	Aggregate Size	Volume Size	Operating System Mount Point
Controller 1	Aggr1	-	6.91TB	-	-
		/omnisci_image_data	-	2.3TB	/omnisci_image_data
		/omnisci_varlib	-	1.4TB	/var/lib/omnisci
		/omnisci_opt	-	150GB	/opt/omnisci
		/omnisci_log	-	425GB	/var/lib/omnisci/data/mapd_log

The `/omnisci_image_data` folder has the sample ships demonstration data that OmniSci provides for validation. The files are compressed with the .gz format, which can be loaded into OmniSci in bulk. The `/var/lib/omnisci` folder has the imported data that is stored in the OmniSci database. The `/opt/omnisci` folder has the OmniSci binary files, and the OmniSci log files are stored in `/var/lib/omnisci/data/mapd_log`.

6 Test Procedure and Detailed Results

6.1 OmniSci Validation in ONTAP AI

To validate OmniSci with NetApp ONTAP AI storage, complete the following steps:

1. Install the OmniSci software according to the instructions.
2. During installation, in the configuration section, enter the environmental variables `OMNISCI_USER`, `OMNISCI_GROUP`, `OMNISCI_STORAGE`, `OMNISCI_PATH`, `OMNISCI_LOG` in `.bashrc` of the root user.

```
# User specific aliases and functions
export OMNISCI_USER=omnisci
export OMNISCI_GROUP=omnisci
export OMNISCI_STORAGE=/var/lib/omnisci
export OMNISCI_PATH=/opt/omnisci
export OMNISCI_LOG=/var/lib/omnisci/data/mapd_log
```

3. Mount the NetApp volumes in the `OMNISCI_STORAGE`, `OMNISCI_PATH`, and `OMNISCI_LOG` folders.

```
root@dgx1-1:~# df -h
Filesystem                          Size  Used Avail Use% Mounted on
udev                                252G     0  252G   0% /dev
tmpfs                                51G  2.3M   51G   1% /run
/dev/sda2                           439G   50G  367G  12% /
tmpfs                               252G     0  252G   0% /dev/shm
tmpfs                               5.0M     0  5.0M   0% /run/lock
tmpfs                               252G     0  252G   0% /sys/fs/cgroup
/dev/sda1                           487M  6.1M  481M   2% /boot/efi
192.168.11.11:/omnisci_image_data   2.3T  518G  1.8T  23% /omnisci_image_data
192.168.11.11:/omnisci_varlib       1.4T  1.2T  258G  82% /var/lib/omnisci
192.168.11.12:/omnisci_opt          150G  421M  150G   1% /opt/omnisci
192.168.11.13:/omnisci_log          425G  126G  300G  30% /var/lib/omnisci/data/mapd_log
tmpfs                                51G     0   51G   0% /run/user/0
root@dgx1-1:~#
```

4. Run the OmniSci initialization process.

```
cd $OMNISCI_PATH/systemd
sudo ./install_omnisci_systemd.sh
```

5. Start `omnisci_server` and `omnisci_web_server` and enable them to start when the system reboots.

```
cd $OMNISCI_PATH
sudo systemctl start omnisci_server
sudo systemctl start omnisci_web_server
sudo systemctl enable omnisci_server
sudo systemctl enable omnisci_web_server
```

6. Check that port 6273 for the OmniSci Immerse visualization system is listening to use the OmniSci portal.

```
root@dgx1-1:/opt/omnisci# netstat -lntp | grep 6273
tcp6       0      0 :::6273                 :::*                    LISTEN
53250/OmniSci_web_s
root@dgx1-1:/opt/omnisci#
```

7. The OmniSci team provides the license for our validation. Update it in the textbox when you connect the OmniSci Immerse visualization system to `http://10.61.218.151:6273/`.

8. In the OmniSci database, create and load the table for the sample ships demonstration data.

```
root@dgx1-1:/var/lib/omnisci# $OMNISCI_PATH/bin/omnisql
Password:
User mapd connected to database mapd
omnisql> CREATE TABLE ships_float(Longitude FLOAT,Latitude FLOAT,unused_z SMALLINT,SOG
DECIMAL(6,2),COG DECIMAL(6,2),Heading SMALLINT,ROT SMALLINT,BaseDateTime TIMESTAMP ENCODING
FIXED(32),Status SMALLINT,VoyageID TEXT ENCODING DICT(32),MMSI TEXT ENCODING
DICT(32),ReceiverType TEXT ENCODING DICT(8),ReceiverID TEXT ENCODING DICT(32),Destination TEXT
ENCODING DICT(32),Cargo TEXT ENCODING DICT(8),Draught SMALLINT,ETA TIMESTAMP ENCODING
FIXED(32),StartTime TIMESTAMP ENCODING FIXED(32),EndTime TIMESTAMP ENCODING
FIXED(32),unused_IMO TIMESTAMP(0),unused_CallSign TEXT ENCODING DICT(16),unused_Name TEXT
ENCODING DICT(8),VesselType TEXT ENCODING DICT (8),VesselLength SMALLINT,Width
SMALLINT,unused_DimensionComponents TEXT ENCODING DICT(32));
```

Note: OmniSci provides the password for the validation.

9. Load the ships sample data.

```
omnisql> copy ships_float from '/omnisci_image_data/omnisci/*_2010*.csv.joined.csv.gz';
Result
Loaded: 1918198950 recs, Rejected: 0 recs in 1406.545000 secs
omnisql> copy ships_float from '/omnisci_image_data/omnisci/*_2011*.csv.joined.csv.gz';
Result
Loaded: 2025721997 recs, Rejected: 0 recs in 1359.854000 secs
omnisql> copy ships_float from '/omnisci_image_data/omnisci/*_2012*.csv.joined.csv.gz';
Result
Loaded: 2086327691 recs, Rejected: 0 recs in 1633.710000 secs
```

```
omnisql>
omnisql> copy ships_float from '/omnisci_image_data/omnisci/*_2013*.csv.joined.csv.gz';
Result
Loaded: 2231356593 recs, Rejected: 0 recs in 1743.824000 secs
omnisql>
omnisql> copy ships_float from '/omnisci_image_data/omnisci/*_2014*.csv.joined.csv.gz';
Result
Loaded: 1835070935 recs, Rejected: 0 recs in 1358.459000 secs
omnisql>
```

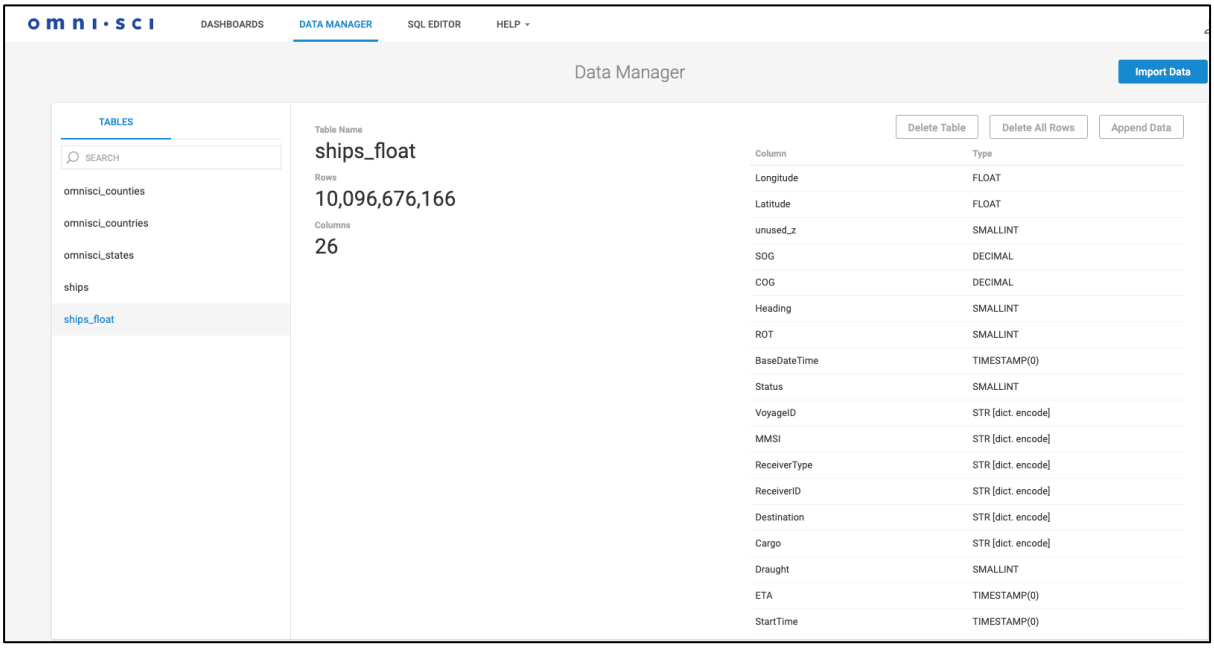
10. Review the following text block, which shows GPU memory use as data is loaded into OmniSci.

```
root@dgx1-1:/var/lib/OmniSci/data/mapd_log# nvidia-smi -l
Tue Jun 25 15:00:30 2019
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 410.104      Driver Version: 410.104      CUDA Version: 10.0     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  Tesla V100-SXM2...  On   | 00000000:06:00.0 Off |                    0 |
| N/A   38C    P0    64W / 300W |  23267MiB / 32480MiB |      0%      Default |
+-------------------------------+----------------------+----------------------+
|   1  Tesla V100-SXM2...  On   | 00000000:07:00.0 Off |                    0 |
| N/A   39C    P0    61W / 300W |  22979MiB / 32480MiB |      0%      Default |
+-------------------------------+----------------------+----------------------+
|   2  Tesla V100-SXM2...  On   | 00000000:0A:00.0 Off |                    0 |
| N/A   40C    P0    57W / 300W |  22979MiB / 32480MiB |      0%      Default |
+-------------------------------+----------------------+----------------------+
|   3  Tesla V100-SXM2...  On   | 00000000:0B:00.0 Off |                    0 |
| N/A   37C    P0    58W / 300W |  20931MiB / 32480MiB |      0%      Default |
+-------------------------------+----------------------+----------------------+
|   4  Tesla V100-SXM2...  On   | 00000000:85:00.0 Off |                    0 |
| N/A   38C    P0    58W / 300W |  20931MiB / 32480MiB |      0%      Default |
+-------------------------------+----------------------+----------------------+
|   5  Tesla V100-SXM2...  On   | 00000000:86:00.0 Off |                    0 |
| N/A   40C    P0    66W / 300W |  22979MiB / 32480MiB |      8%      Default |
+-------------------------------+----------------------+----------------------+
|   6  Tesla V100-SXM2...  On   | 00000000:89:00.0 Off |                    0 |
| N/A   41C    P0    63W / 300W |  20931MiB / 32480MiB |      0%      Default |
+-------------------------------+----------------------+----------------------+
|   7  Tesla V100-SXM2...  On   | 00000000:8A:00.0 Off |                    0 |
| N/A   39C    P0    59W / 300W |  22979MiB / 32480MiB |      8%      Default |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                       GPU Memory |
|  GPU       PID   Type   Process name                             Usage      |
|=============================================================================|
|    0     52883    C+G   /opt/omnisci/bin/omnisci_server            23251MiB |
|    1     52883    C+G   /opt/omnisci/bin/omnisci_server            22963MiB |
|    2     52883    C+G   /opt/omnisci/bin/omnisci_server            22963MiB |
|    3     52883    C+G   /opt/omnisci/bin/omnisci_server            20915MiB |
|    4     52883    C+G   /opt/omnisci/bin/omnisci_server            20915MiB |
|    5     52883    C+G   /opt/omnisci/bin/omnisci_server            22963MiB |
|    6     52883    C+G   /opt/omnisci/bin/omnisci_server            20915MiB |
|    7     52883    C+G   /opt/omnisci/bin/omnisci_server            22963MiB |
+-----------------------------------------------------------------------------+
```

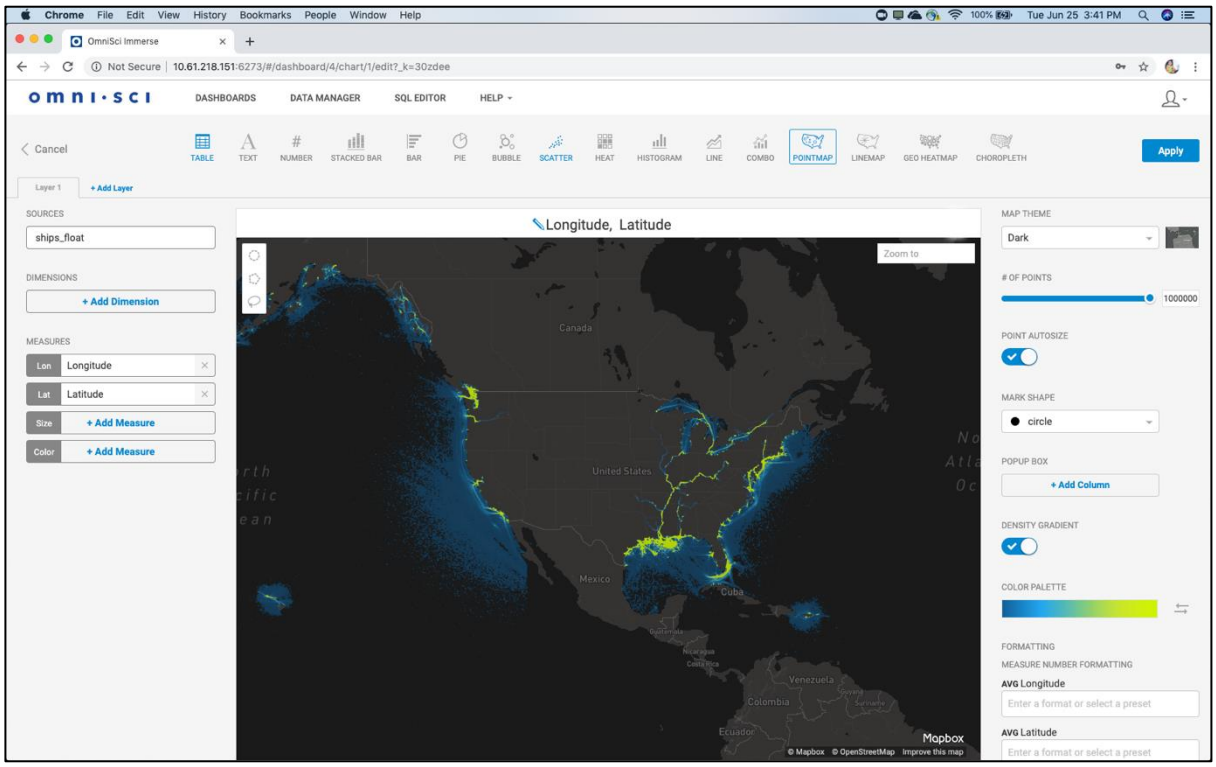
11. Review the following text block, which shows the OmniSci configuration that is used in the validation.

```
root@dgx1-1:/var/lib/omnisci/data/mapd_log# cd /var/lib/omnisci/
root@dgx1-1:/var/lib/omnisci# ls -ltrha  omnisci.conf
-rw-r--r-- 1 root root 188 Jun 20 18:02 omnisci.conf
root@dgx1-1:/var/lib/omnisci# cat omnisci.conf
port = 6274
http-port = 6278
calcite-port = 6279
data = "/var/lib/omnisci/data"
null-div-by-zero = true
render-mem-bytes = 2000000000

[web]
port = 6273
frontend = "/opt/omnisci/frontend"
root@dgx1-1:/var/lib/omnisci#
```

12. Verify that the table `ships_float` was created in the Data Manager section.



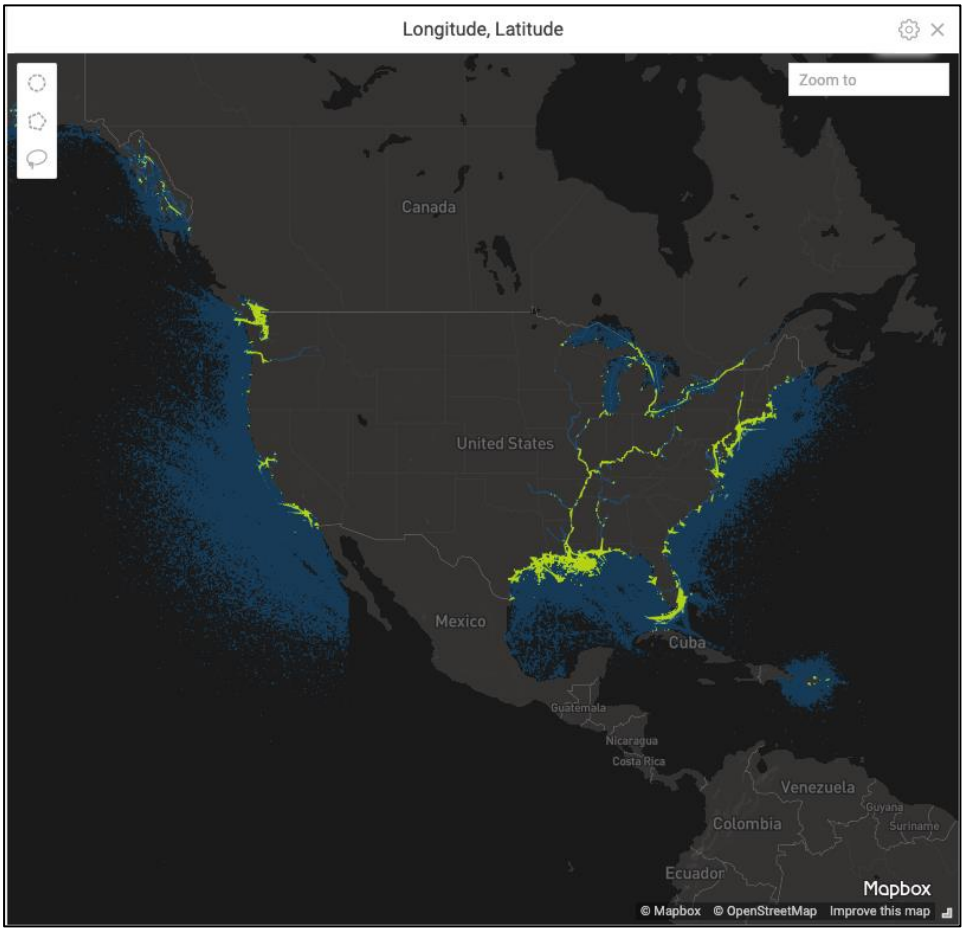
13. To create a new dashboard and point map and to verify that the background rendering is working, complete the following steps:



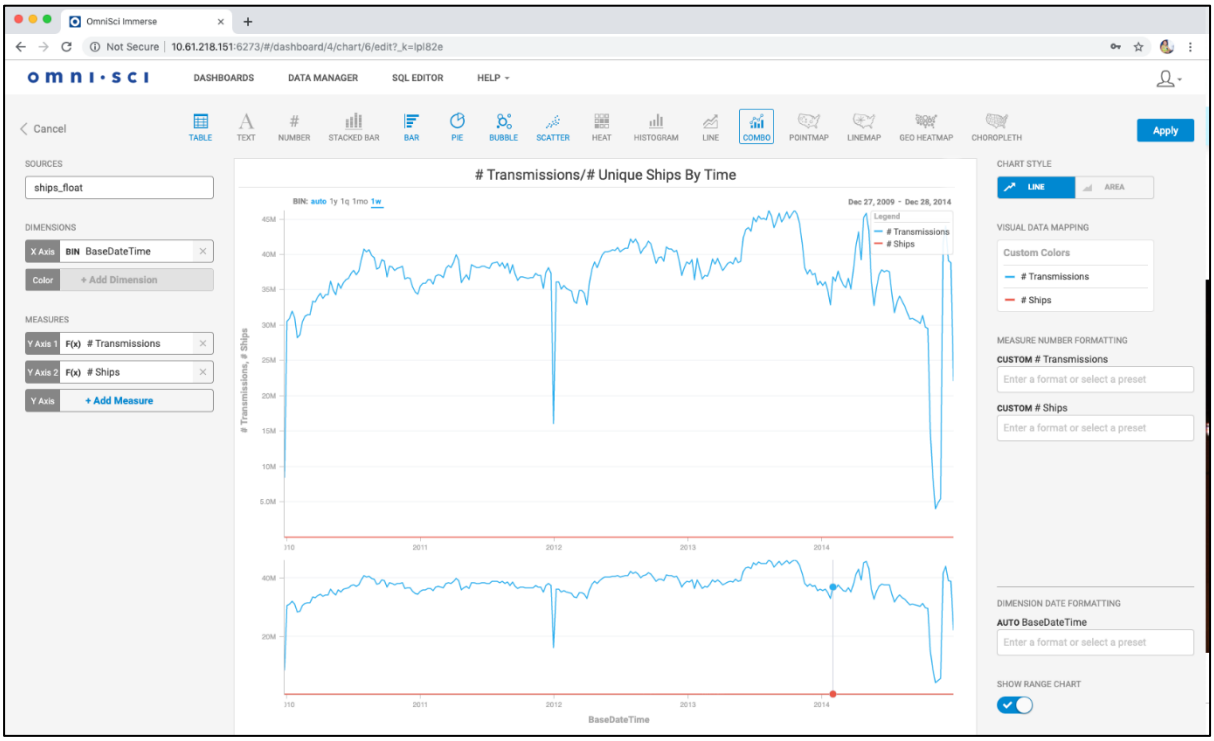
a. Click New Dashboard and enter the title in Dashboard Title Required.
b. Click Add Chart and then select Pointmap.
c. Click Sources and select ships_float.
d. Click Lon and select Longitude.
e. Click Lat and select Latitude.
f. Select Dark from Map Theme.

g. Enable Point Autosize and select Circle from Mark Shape.

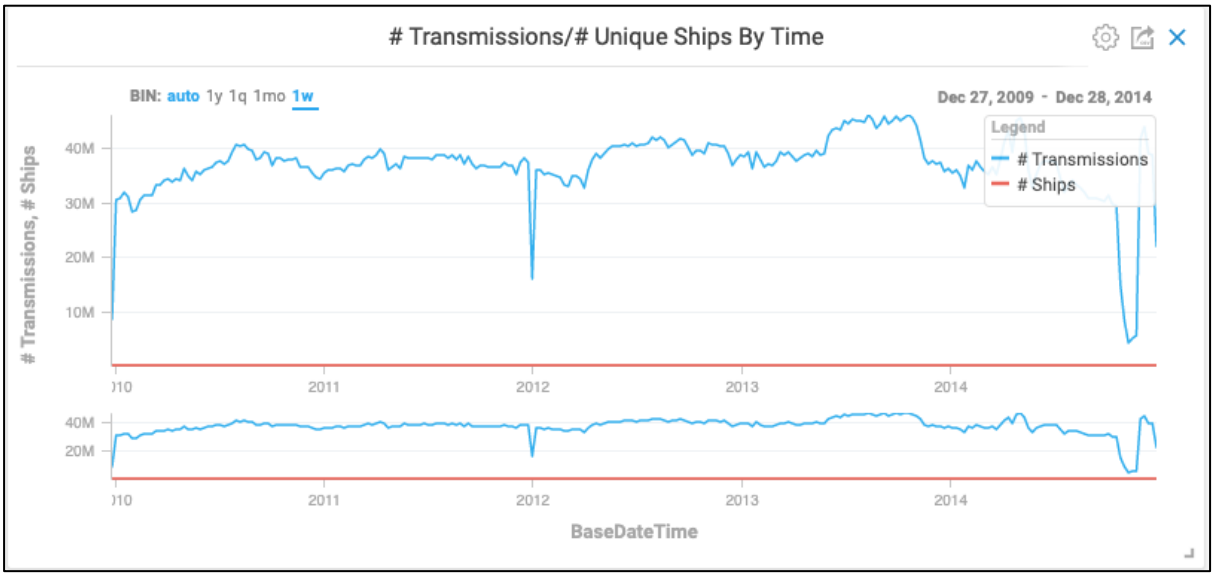
h. Click Apply to finish.



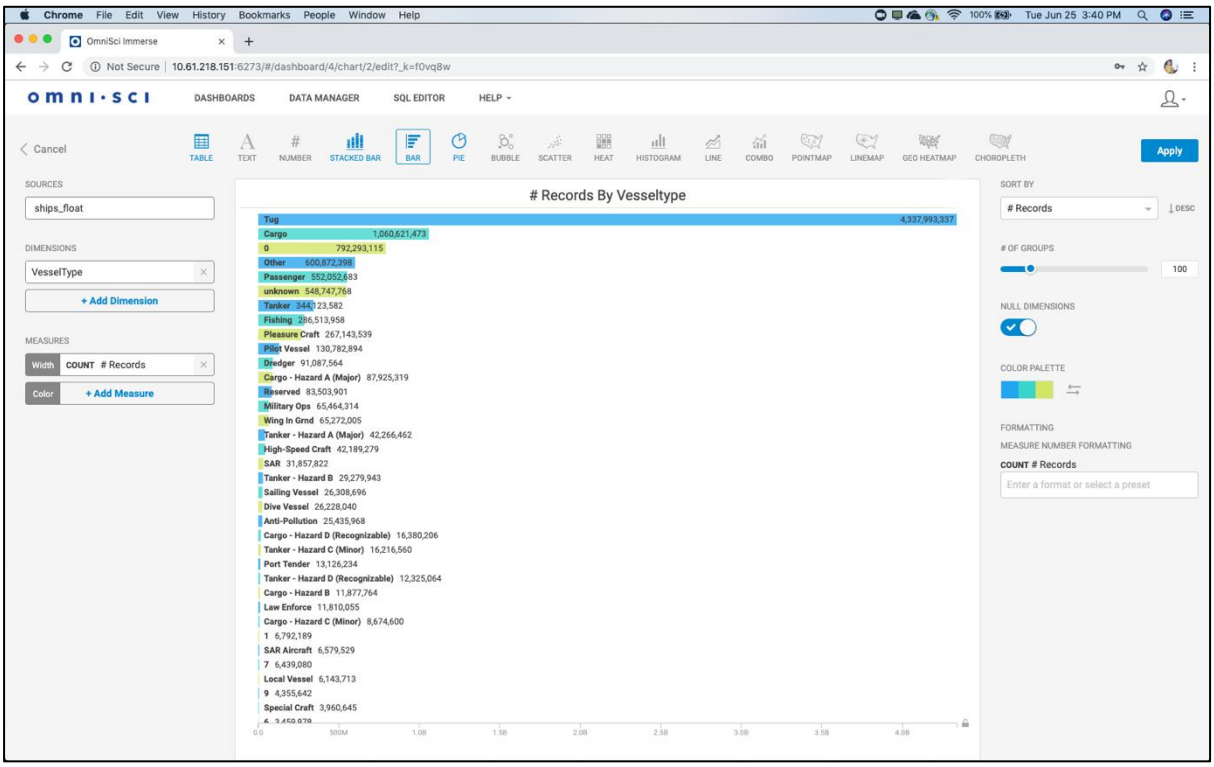
14. The following steps create a combo chart. To find the transmissions per number of unique ships by time, complete the following steps:



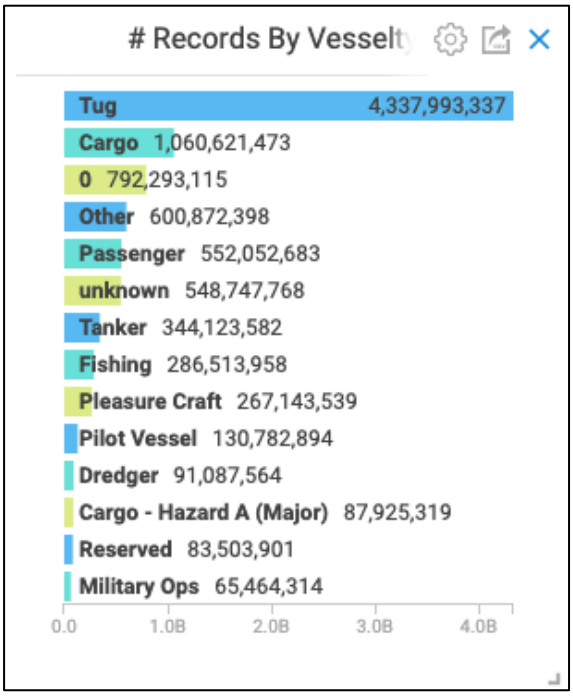
a. Click Add Chart and then click Combo.
b. Click Sources and select ships_float.
c. Select BaseDateTime for the X Axis (Binning – Bin Unit and select Auto).
d. For Y Axis 1 F(x), select:
    i. # Transmissions
    ii. Custom Measure: count(*)
e. For Y Axis 2 F(x), select:
    i. Name: # Ships
    ii. Custom Measure: approx._count_distinct(MMSI)
f. Select Line for the Chart Style.
g. Enable Show Range Chart.
h. Select Dark for the Map Theme.
i. Enable Point Autosize and select Circle from Mark Shape.
j. Click Apply.



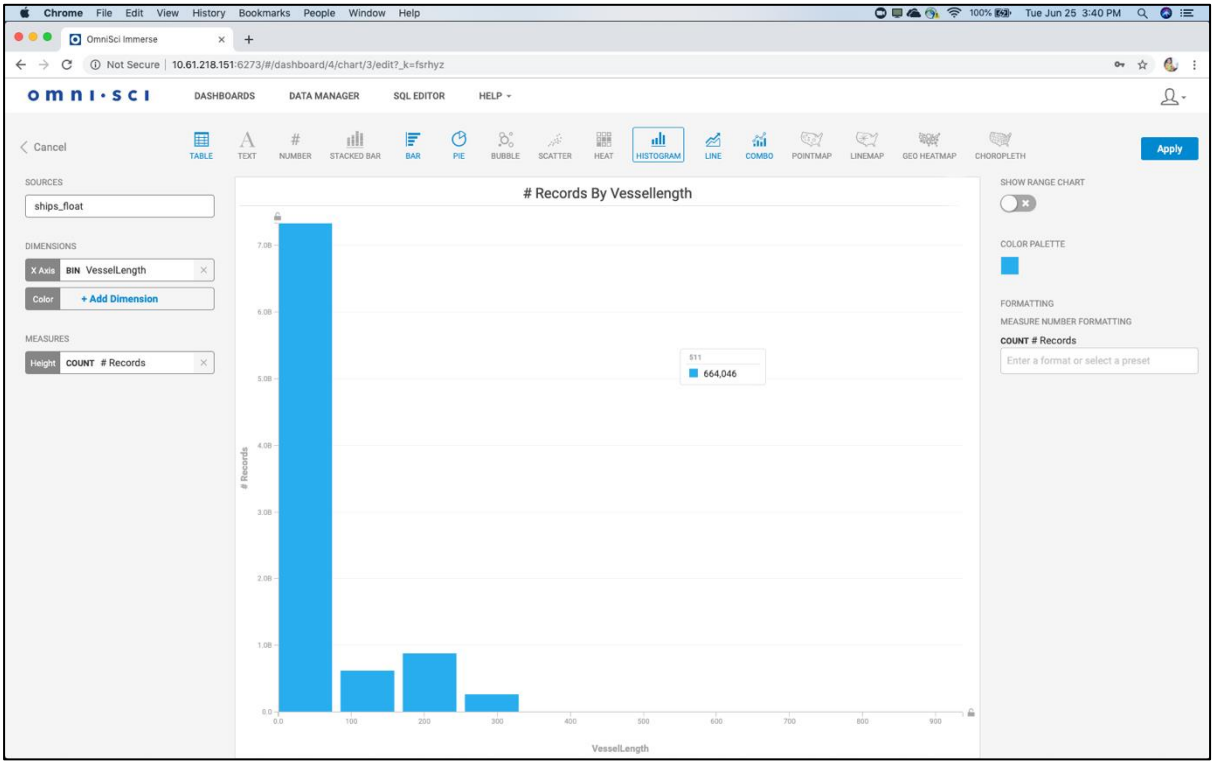
15. The following steps create a bar chart. To find the number of records by vessel type, complete the following steps:



a. Click Add Chart and then click Bar.
b. Click Sources and select ships_float.
c. Select VesselType from Dimensions.
d. Under Measures, click Width and select # Records from the drop-down list.
e. Under Sort By, select # Records.
f. Enable Null Dimensions.
g. Click Apply.



16. The following steps create a histogram chart. To find the number of records by the vessel length, complete the following steps:

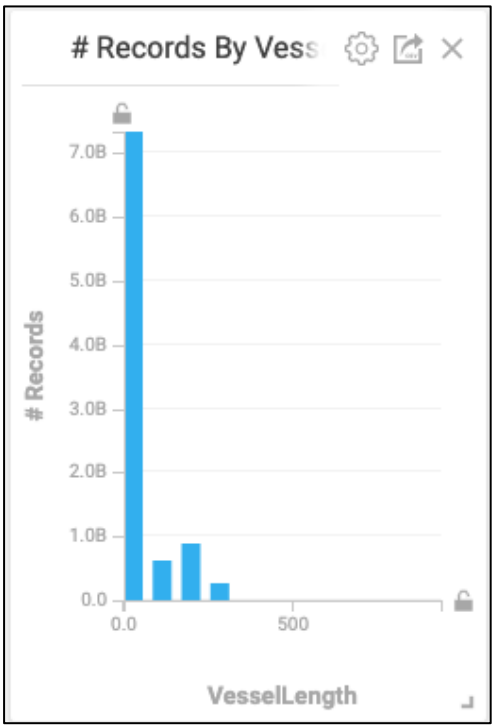


a. Click Add Chart and then click Histogram.
b. Click Sources and select ships_float.
c. For X Axis, select VesselLength:
    i. Enable Binning: ON
    ii. # of BINS: 12
d. Under Measures, click Height and select # Records.

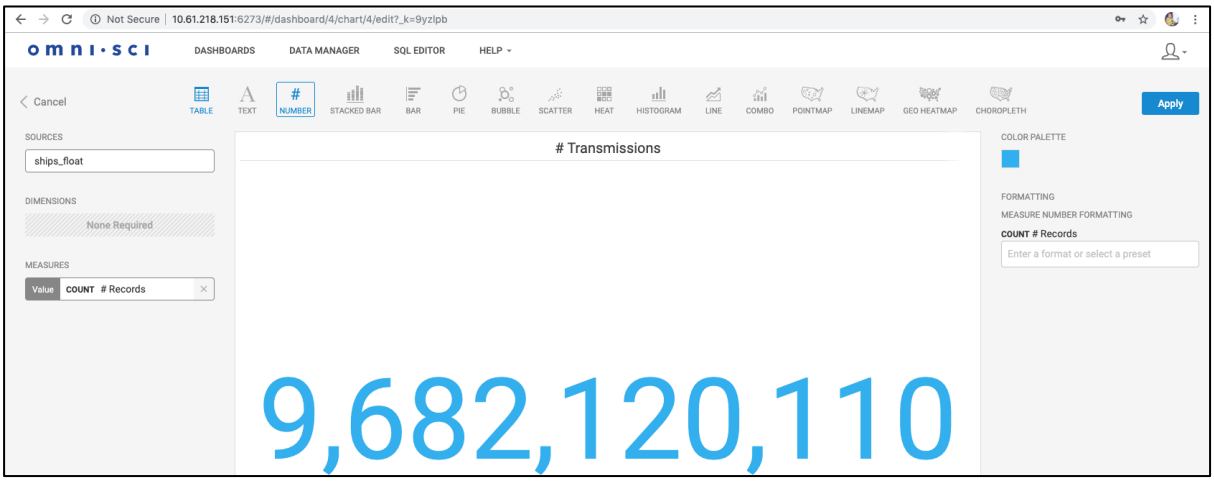
e. Disable Show Range Chart.

f. Click Apply.

Note: We included `bigint-count = true` in `/var/lib/OmniSci/OmniSci.conf` to get the correct chart.



17. The following steps create a number chart. To show the number of transmissions and the number of unique ships, complete the following steps:

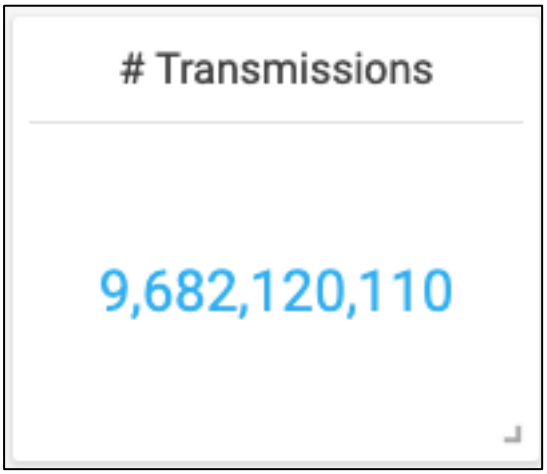


a. Click Add Chart and then click Number.

b. Click Sources and select ships_float.

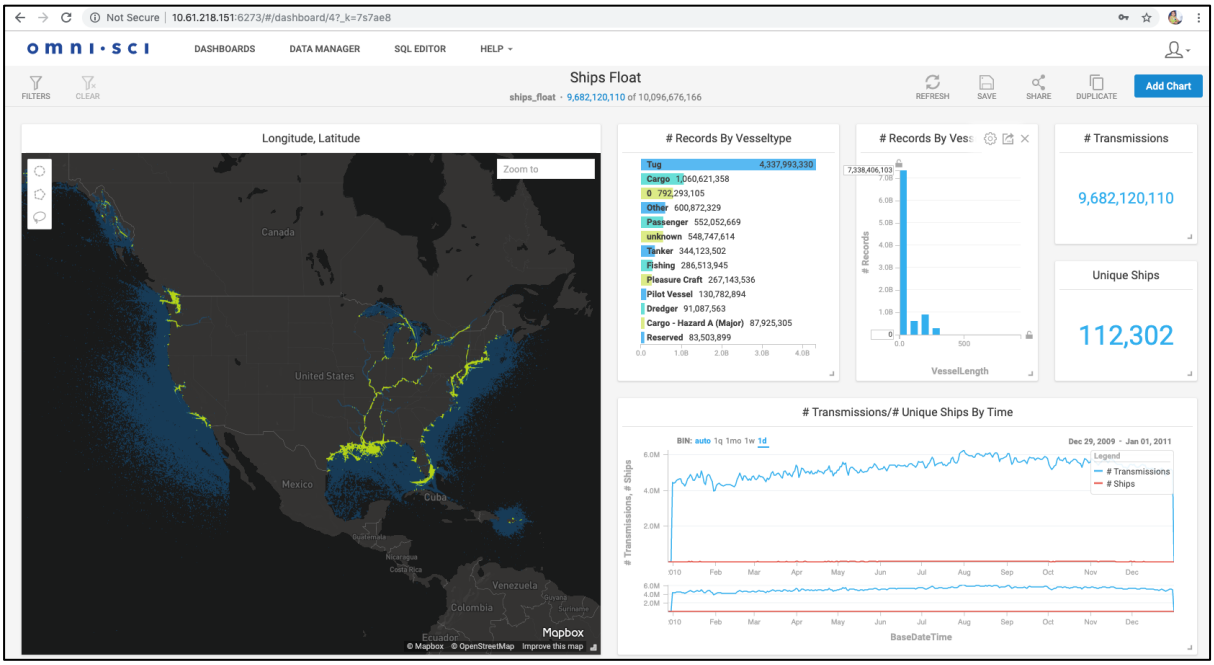
c. Under Measures, click Value and select # Records from the drop-down list.

d. Click Apply.



Note: We can also find the number of ships by following the same steps.

The following screenshot shows all the charts in the same window, which helps you visualize a higher level of the ships data.



6.2 ONTAP Storage Efficiency Test

After we completed the data loading and visualization, we queried storage efficiency by using the ONTAP command line to determine the following parameters:

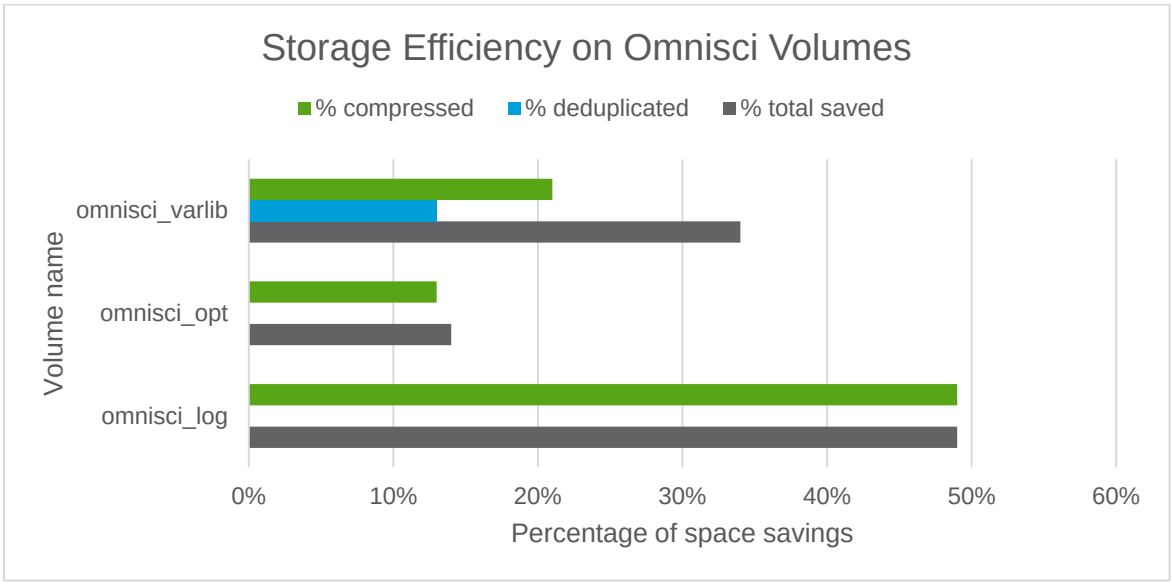
- Total space savings as a percentage due to ONTAP inline storage efficiency.
- Percentage of compressed data and deduplicated data.

Based on our validation with the ships data, we observed the following (Figure 3):

- With OmniSci, we found savings from a combination of volume deduplication and volume compression.
- For `omnisci_varlib` (OmniSci data storage), 21% of the savings was from ONTAP compression, and 13% of the savings was from ONTAP deduplication, for a total storage savings of 34%.
- For `omnisci_opt` (OmniSci binary), 13% of the savings was from ONTAP compression and 0.5% of the savings was from ONTAP deduplication, for a total storage savings of 14%.

- For `omnisci_log` (OmniSci log files), 49% of the savings was from ONTAP compression, and 0% of the savings was from ONTAP deduplication, for a total storage savings of 49%.

Figure 3) Storage efficiency with ONTAP AI and OmniSci.



The logical size of the OmniSci data was approximately 2TB, but the actual physical space that was used was only about 1.6TB. The space savings is the result of ONTAP inline storage efficiency. Although results can vary based on actual data, the results that we obtained were typical across all the AFF A800 configurations that we tested.

6.3 OmniSci Cloning by Using NetApp FlexClone

The purpose of this validation was to demonstrate the use of NetApp FlexClone technology to create fast, space-efficient copies of production data. We performed the following procedure:

1. The source OmniSci ran in `10.61.218.151` (the host name was `dgx1-1`), and the cloned OmniSci ran in `10.61.218.152` (the host name was `dgx1-2`).
2. We created a Snapshot copy of OmniSci volumes, including binary volumes, data volumes, and log file volumes, with a consistency group (CG) Snapshot copy. We used ONTAP software development kit (SDK) API calls to create CG Snapshot copies with Python.

```
root@dgx1-1:/usr/src/netapp-manageability-sdk-9.6/src/sample/Data_ONTAP/Python# pwd
/usr/src/netapp-manageability-sdk-9.6/src/sample/Data_ONTAP/Python

root@dgx1-1:/usr/src/netapp-manageability-sdk-9.6/src/sample/Data_ONTAP/Python# python
cg_operation.py 10.61.218.100 admin cpocai cg-start omnicgstart relaxed onmisci_varlib
onmisci_opt onmisci_log
Consistency Group operation started successfully with cg-id=600975308

root@dgx1-1:/usr/src/netapp-manageability-sdk-9.6/src/sample/Data_ONTAP/Python# python
cg_operation.py 10.61.218.100 admin cpocai cg-commit 600975308 relaxed onmisci_varlib
onmisci_opt onmisci_log
Consistency Group operation commited successfully

root@dgx1-1:/usr/src/netapp-manageability-sdk-9.6/src/sample/Data_ONTAP/Python#


wopr::> snapshot list -vserver cpocai_nfs -volume omni* -snapshot omnicgstart
                                                                  ---Blocks---
Vserver  Volume   Snapshot                                  Size Total% Used%
-------- -------- ------------------------------------ -------- ------ -----
cpocai_nfs
         onmisci_log
                  omnicgstart                              168KB     0%    0%
         onmisci_opt
                  omnicgstart                              152KB     0%    0%
         onmisci_varlib
```

```
                  omnicgstart                                 312KB     0%    0%
3 entries were displayed.

wopr::>
```

3. We created FlexClone volumes of the OmniSci data volume, binary volume, and log file volume.

```
wopr::> volume clone create -parent-snapshot omnicgstart -flexclone onmisci_log_cl -type RW -
parent-volume onmisci_log
[Job 3031] Job succeeded: Successful

wopr::> volume clone create -parent-snapshot omnicgstart -flexclone onmisci_opt_cl -type RW -
parent-volume onmisci_opt
[Job 3032] Job succeeded: Successful

wopr::> volume clone create -parent-snapshot omnicgstart -flexclone onmisci_varlib_cl -type RW
-parent-volume onmisci_varlib
[Job 3033] Job succeeded: Successful
```

4. We mounted the volume in the ONTAP storage controller.

```
wopr::> mount -volume onmisci_opt_cl -junction-path /onmisci_opt_cl
  (volume mount)

wopr::> mount -volume onmisci_log_cl -junction-path /onmisci_log_cl
  (volume mount)

wopr::>
wopr::> mount -volume onmisci_varlib_cl -junction-path /onmisci_varlib_cl
  (volume mount)
```

5. We identified the export path (junction path) of the FlexClone volumes in the NetApp storage controller.

```
wopr::> volume show -vserver cpocai_nfs -volume onmisci_*_cl -fields junction-path
vserver    volume         junction-path
---------- -------------- ---------------
cpocai_nfs onmisci_log_cl /onmisci_log_cl
cpocai_nfs onmisci_opt_cl /onmisci_opt_cl
cpocai_nfs onmisci_varlib_cl
                          /onmisci_varlib_cl
3 entries were displayed.
```

6. We updated `/etc/fstab`, created the required folders, and mounted the FlexClone volumes in the operating system in another DGX-1 server (`dgx1-2`).

```
root@dgx1-2:~# cat /etc/fstab
# /etc/fstab: static file system information.
#
# Use 'blkid' to print the universally unique identifier for a
# device; this may be used with UUID= as a more robust way to name devices
# that works even if disks are added and removed. See fstab(5).
#
# <file system> <mount point>   <type>  <options>       <dump>  <pass>
# / was on /dev/sda2 during installation
UUID=1afcf8f1-db9a-4919-a447-a4f70addafe8 /               ext4    errors=remount-ro 0       1
# /boot/efi was on /dev/sda1 during installation
UUID=6DA7-A131  /boot/efi       vfat    umask=0077      0       1


# Uncomment the next line to mount raid on /raid
#/dev/sdb1 /raid ext4 defaults,nofail 0 2
10.61.218.8:/vol/poc/81908-WOPR /mnt/poc nfs defaults 0 0
10.61.100.20:/vol/data /mnt/data nfs defaults 0 0
192.168.11.12:/pb_fg_all /mnt/mount_0 nfs defaults 0 0
192.168.12.11:/pb_fg_all /mnt/mount_1 nfs defaults 0 0
192.168.11.12:/testvol01 /testvol01 nfs defaults 0 0
#onmisci clone
192.168.11.11:/onmisci_varlib_cl /var/lib/onmisci nfs defaults 0 0
192.168.11.12:/onmisci_opt_cl /opt/onmisci nfs defaults 0 0
192.168.11.13:/onmisci_log_cl /var/lib/onmisci/data/mapd_log nfs defaults 0 0
root@dgx1-2:~#
root@dgx1-2:~# mkdir -p /var/lib/onmisci
root@dgx1-2:~# mkdir -p /opt/onmisci
root@dgx1-2:~# mkdir -p /var/lib/onmisci/data/mapd_log
```

7. We installed the required OmniSci packages in the new DGX-1 server (`dgx1-2`).

```
Sudo apt update
Sudo apt upgrade
Sudo apt install default-jre-headless
Sudo apt install apt-transport-https
sudo useradd -U -m onmisci
sudo apt-get install linux-headers-$(uname -r)
sudo curl -O https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64/cuda-
repo-ubuntu1804_10.0.130-1_amd64.deb
sudo dpkg -i cuda-repo-ubuntu1804_10.0.130-1_amd64.deb
sudo apt-key adv --fetch-keys
http://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86_64/7fa2af80.pu
sudo apt update
sudo apt install cuda-drivers linux-image-extra-virtual
sudo reboot
curl https://releases.onmisci.com/GPG-KEY-onmisci | sudo apt-key add -
echo "deb https://releases.onmisci.com/ee/apt/ stable cuda" | sudo tee
/etc/apt/sources.list.d/onmisci.list
sudo apt update
sudo apt install onmisci
update .bashrc with onmisci variables.
```

8. We removed the existing lock file, which was created in another OmniSci instance.

```
ls -ltrah /var/lib/onmisci/data/
mv onmisci_server_pid.lck onmisci_server_pid.lck_old
```

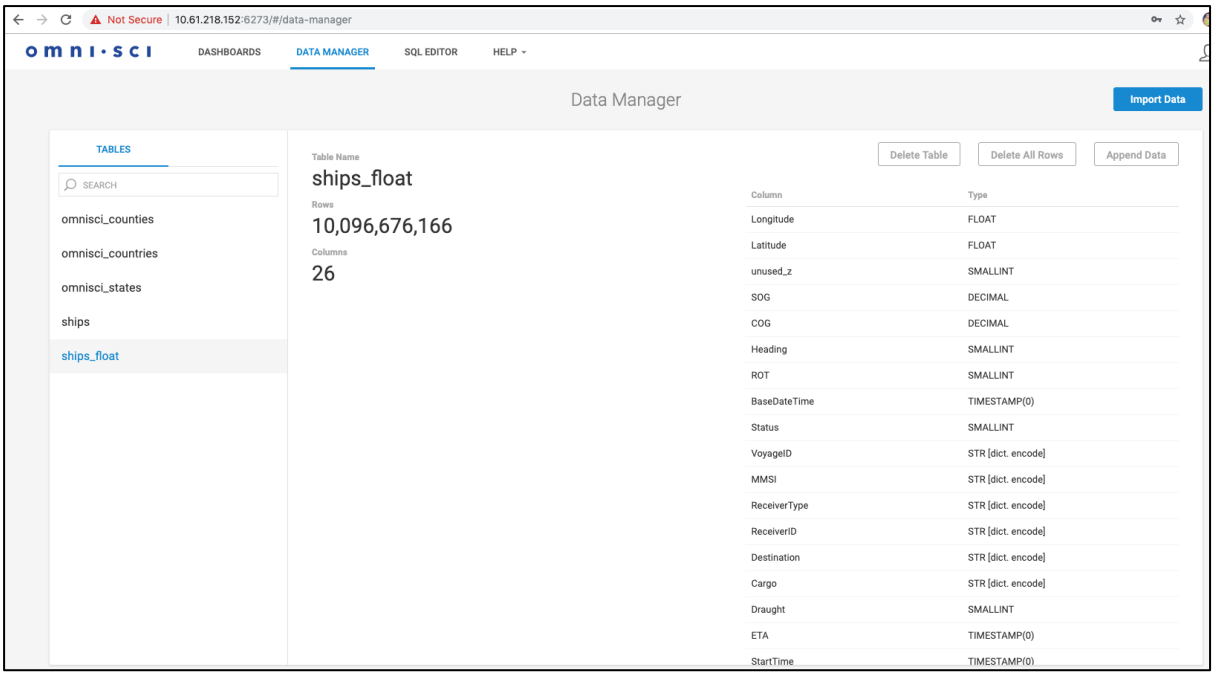
9. We started `omnisci_server` and `omnisci_web_server`.

```
sudo systemctl restart onmisci_server
sudo systemctl enable onmisci_web_server
```

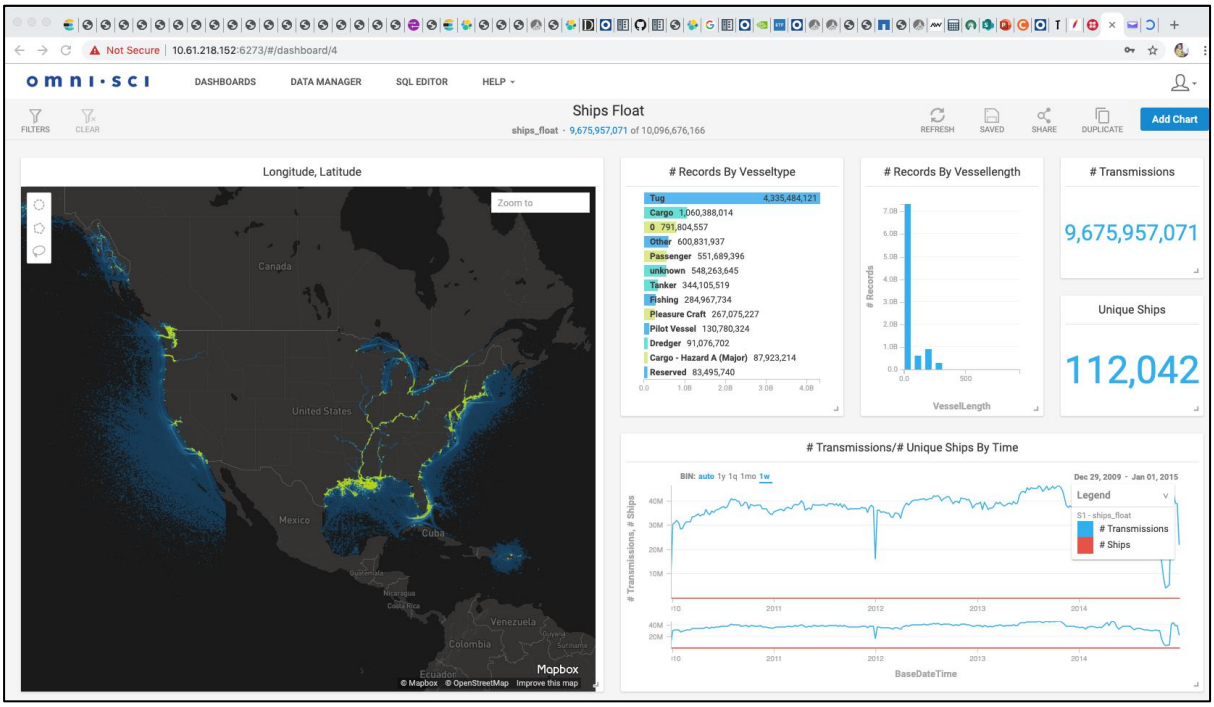
10. We checked the new PID update in `omnisci_server_pid.lck` and checked the `INFO` log.

```
root@dgx1-2:~# ps -efwww | grep -i omni
root     23983 23061  0 17:47 pts/0    00:00:00 grep --color=auto -i omni
root     27143     1  0 Jun26 ?        00:03:32 /opt/onmisci/bin/onmisci_web_server --config
/var/lib/onmisci/onmisci.conf
root     33666     1  3 Jun27 ?        06:43:35 /opt/onmisci/bin/onmisci_server --config
/var/lib/onmisci/onmisci.conf
root     33729 33666  0 Jun27 ?        00:15:18 -Xmx1024m -DMAPD_LOG_DIR=/var/lib/onmisci/data
-jar /opt/onmisci/bin/calcite-1.0-SNAPSHOT-jar-with-dependencies.jar -e
/opt/onmisci/QueryEngine/ -d /var/lib/onmisci/data -p 6279 -m 6274 -T  -P
root@dgx1-2:~# cat /var/lib/onmisci/data//onmisci_server_pid.lck
33666
root@dgx1-2:~#
cd /var/lib/onmisci/data/mapd_log/
tail -f *.INFO
```

11. We checked the number of records that were similar to the source OmniSci server.



12. OmniSci brought the data to the GPU memory to display it. This process might take more time relative to the source OmniSci.



7 Conclusion

This NetApp ONTAP AI and OmniSci solution is ideal for your organization to visualize and to analyze your data with enterprise-level data protection. Some key benefits are:

- Higher performance
- A reduction in storage requirements
- A more robust OmniSci configuration
- Enterprise-class data protection
- Consistent performance, even if storage hardware fails
- Significant savings in storage utilization because of ONTAP inline storage efficiency
- The creation of fast, space-efficient data copies with NetApp FlexClone technology

Acknowledgments

- Ranganathan Sunday, Senior Product Manager, Artificial Intelligence, NetApp
- Santosh Rao, Senior Technical Director, NetApp
- Nilesh Bagad, Senior Product Manager, Artifical Intelligence and Big Data, NetApp
- Rebecca Kao, Director of Global Partner Marketing, OmniSci
- James Yu, Head of Strategic Alliances, OmniSci
- Chris Bithos, Alliance Program Manager, NetApp

Where to Find Additional Information

To learn more about the information that is described in this document, see the following documents and/or websites:

- OmniSci installation and its details
  https://www.OmniSci.com/docs/latest/4_ubuntu-tarball-gpu-ee-recipe.html

- NetApp all-flash arrays product page
  https://www.netapp.com/us/products/storage-systems/all-flash-array/aff-a-series.aspx
- NetApp ONTAP data management software product page
  http://www.netapp.com/us/products/data-management-software/ontap.aspx

Version History

Version	Date	Document Version History
Version 1.0	August 2019	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.