

Storage Capacity Analysis

Nicolas Genibre

September 26, 2016

Introduction

What is the Storage Capacity looks like for the NetApp cluster in Fort Lauderdale Laboratory?

For this Analysis we implemented the NetApp Perfstat tool to run on a schedule, for 16 days, and captured 48 iterations of 15 minutes each day using the following to determine capacity:

df -x -m (pre and prostats); df -A -m (pre and prostats); sysstat_x_1sec ; sysstat_x_5sec

Clean Data

The raw data is composed of 32 files, 16 files per NetApp controller. Each files are plain text output containing the iterations for a day.

We're cleaning the data with the help of shell scripts to transform the data into a csv type file:

aggr.sh; vol.sh and the result is in the data directory as aggr.csv and vol.csv

```
myaggr <- read.csv("../data/aggr.csv")
myvol <- read.csv("../data/vol.csv")
```

The **Time** is expressed in epoch date. There are two **Filers** in the Cluster.

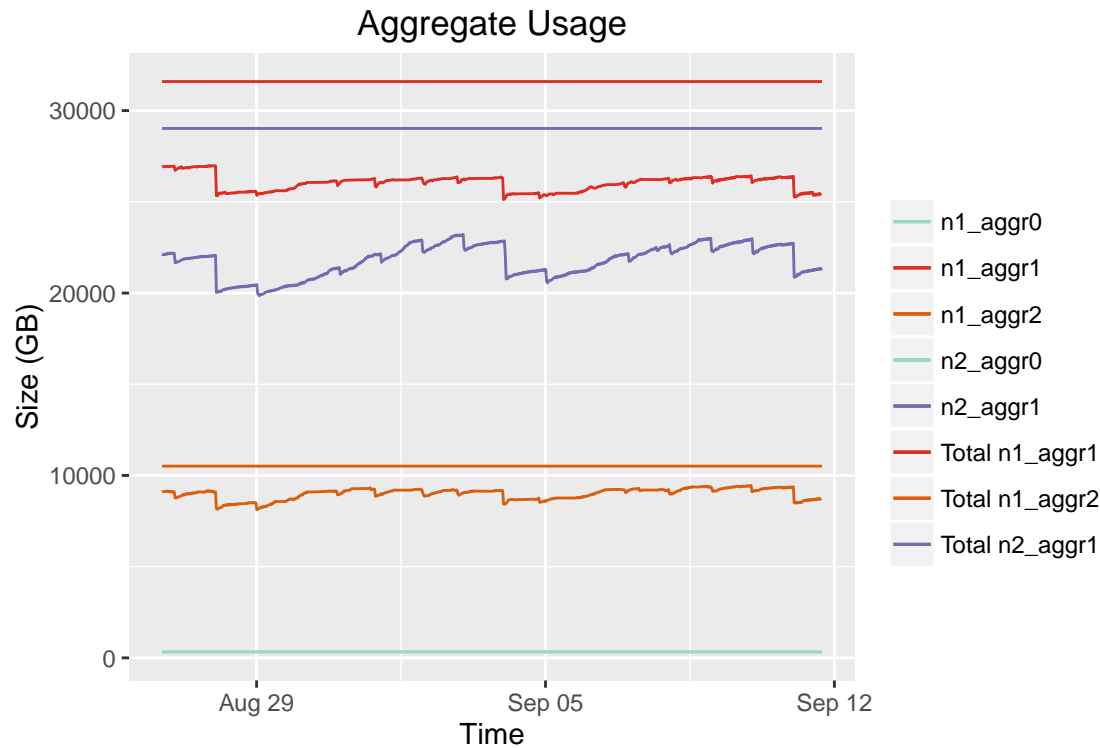
Total, **Used** and **Free** capacity are measurements in MB. The **Ratio** is the expression in percentage of Used versus Total.

Aggregate

```
library(dplyr)
library(ggplot2)
```

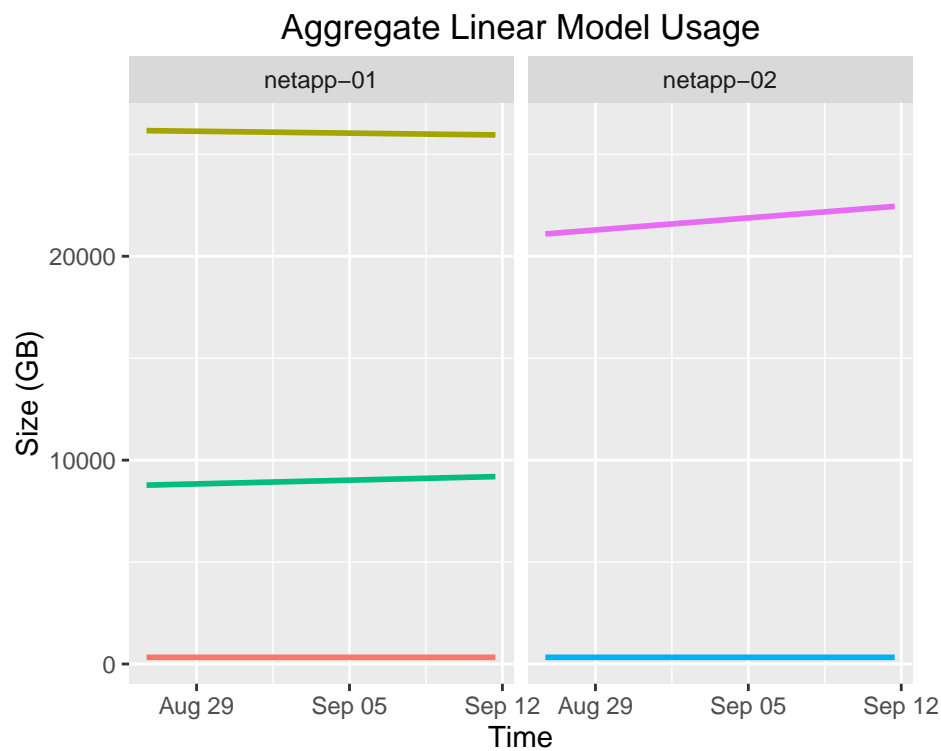
```
head(myaggr)
```

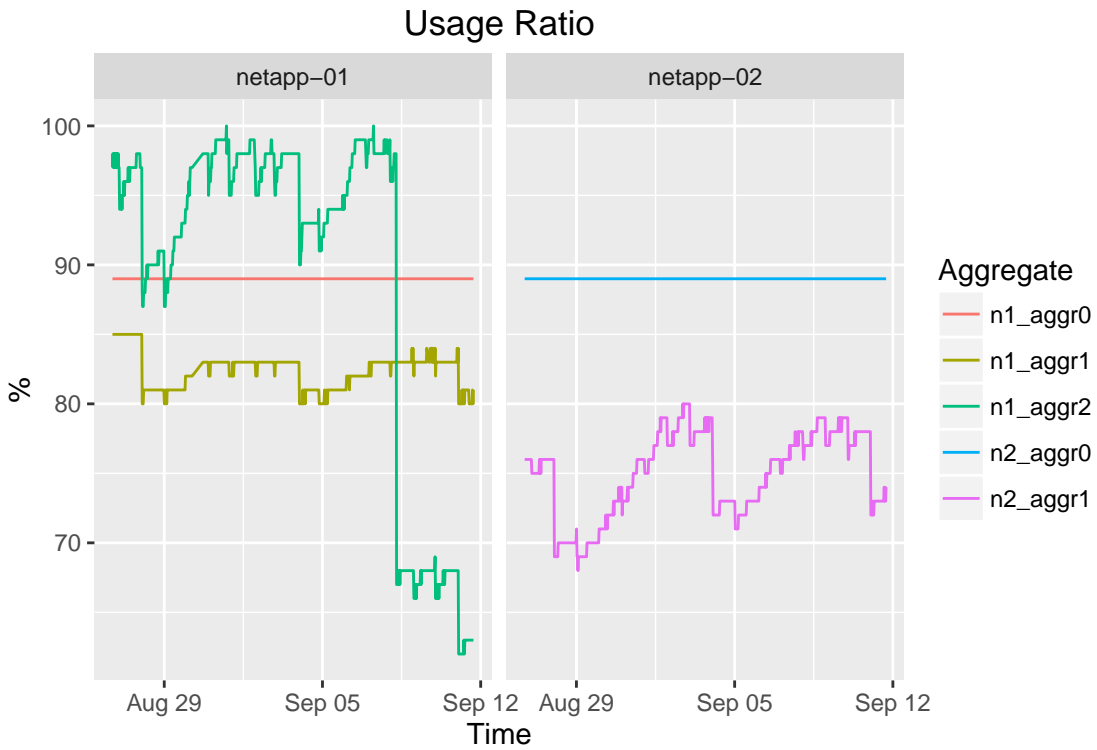
##	Time	Filer	Aggregate	Total	Used	Free	UsageRatio
## 1	1472245275	netapp-02	n2_aggr0	376181	334807	41373	89
## 2	1472245275	netapp-02	n2_aggr1	29718342	22655655	7062686	76
## 3	1472247100	netapp-02	n2_aggr0	376181	334807	41373	89
## 4	1472247100	netapp-02	n2_aggr1	29718342	22620766	7097575	76
## 5	1472248921	netapp-02	n2_aggr0	376181	334807	41373	89
## 6	1472248921	netapp-02	n2_aggr1	29718342	22661474	7056868	76



The Aggregate' used capacity has not change much from the beginning to the end of the data collection.

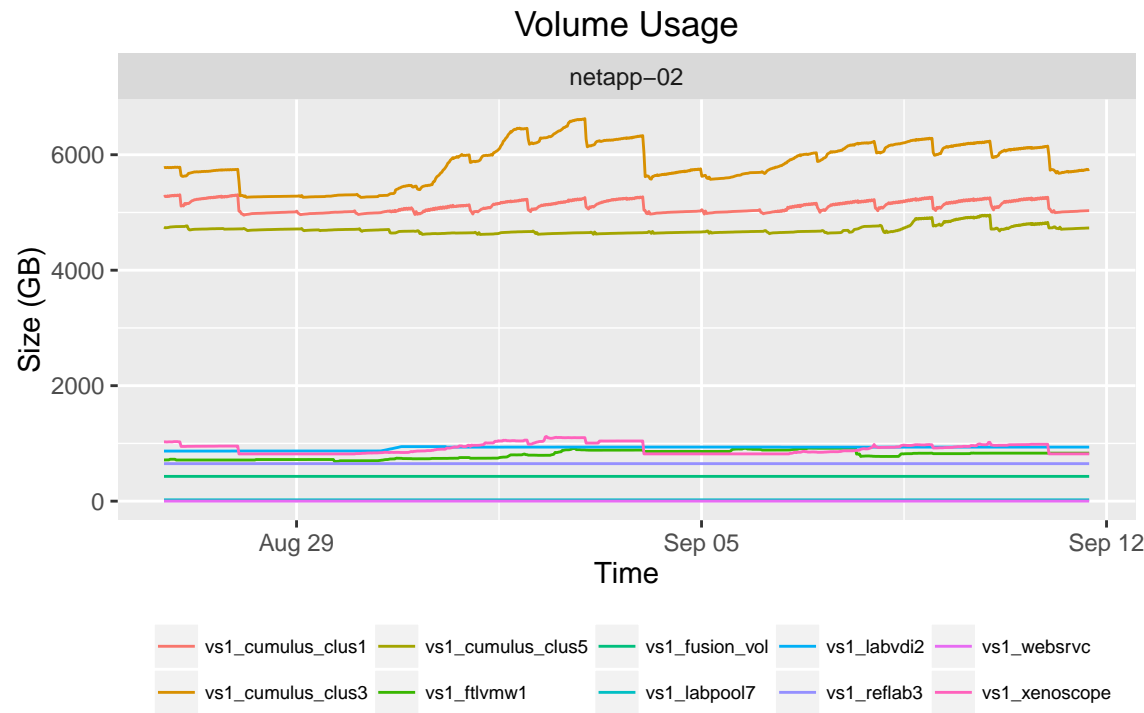
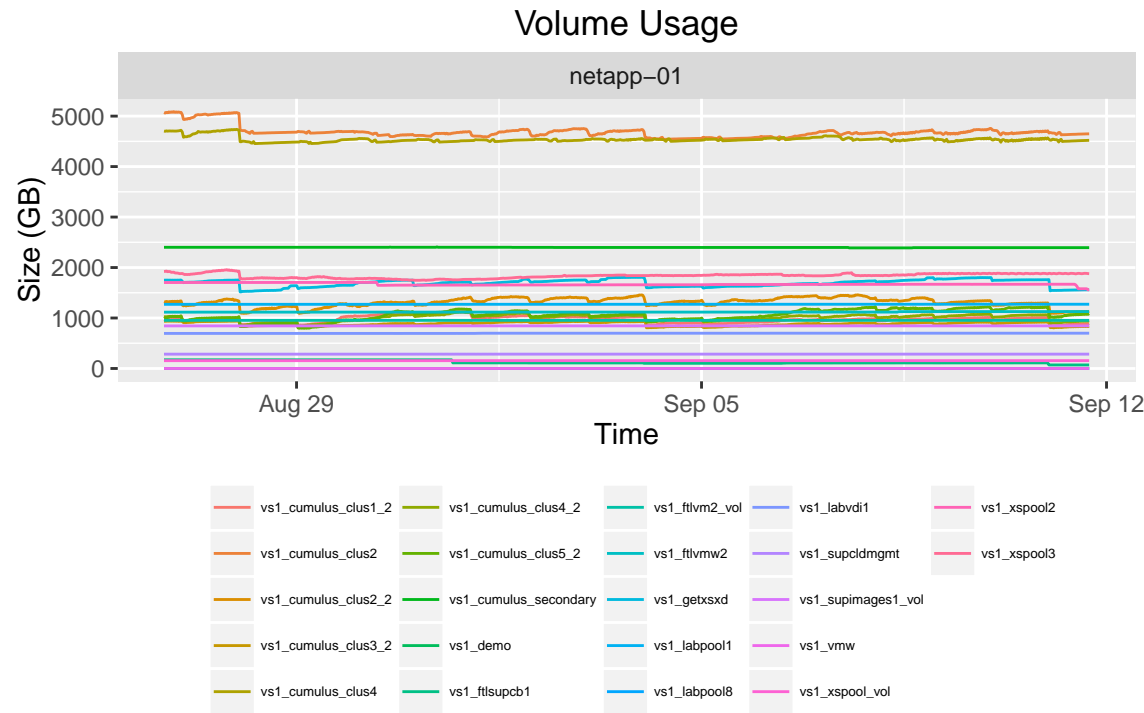
According the linear model we see a slight decrease of Aggregate 1 and increase of Aggregate 2 on Filer 1 over time. On Filer 2 the increase is larger for the Argggregate 2.



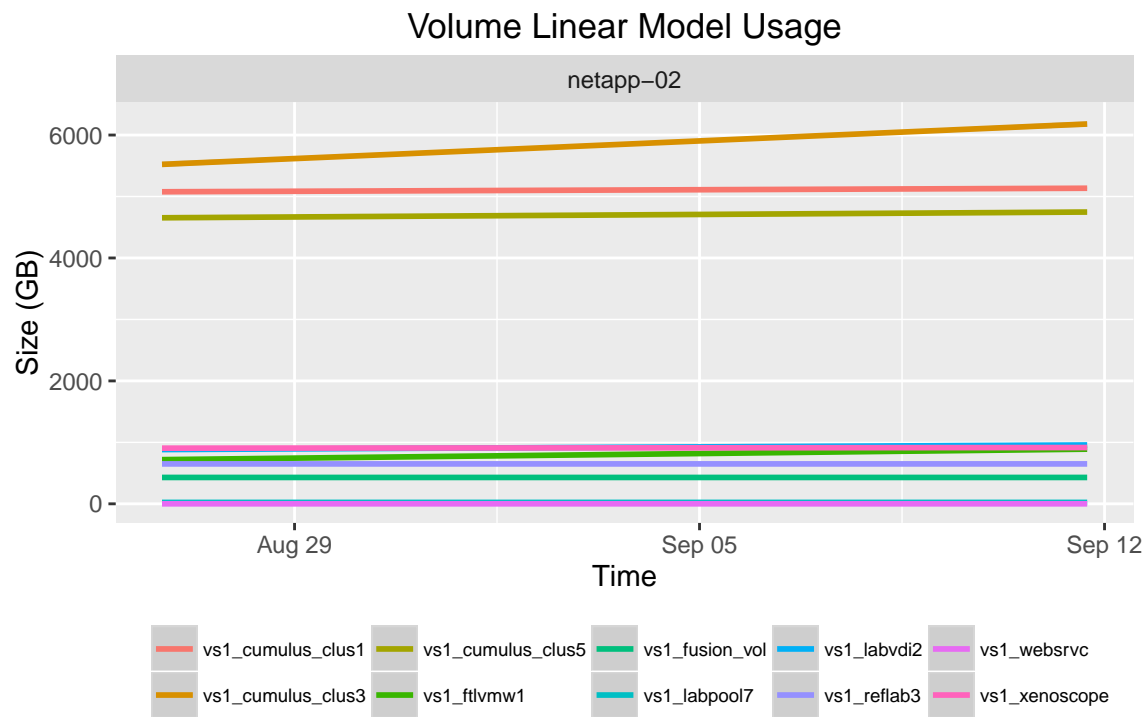
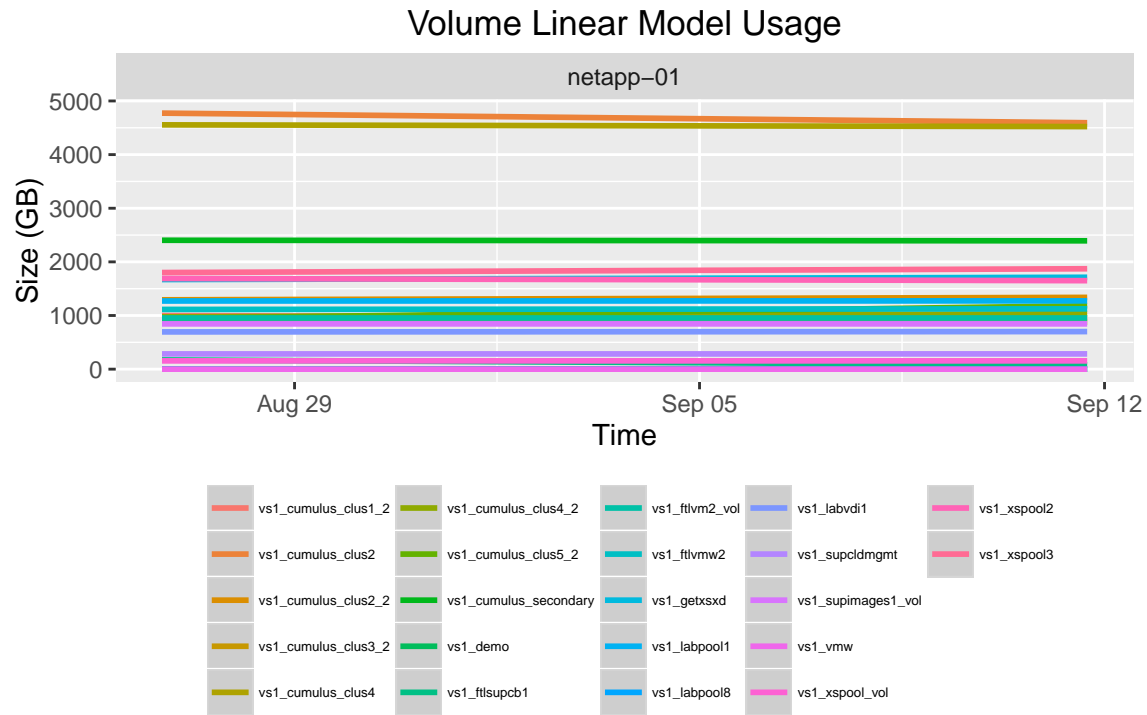


Aggregates are kept under the 85% threshold. During the collection we can see that one of the Aggregate on Filer 1 was increased.

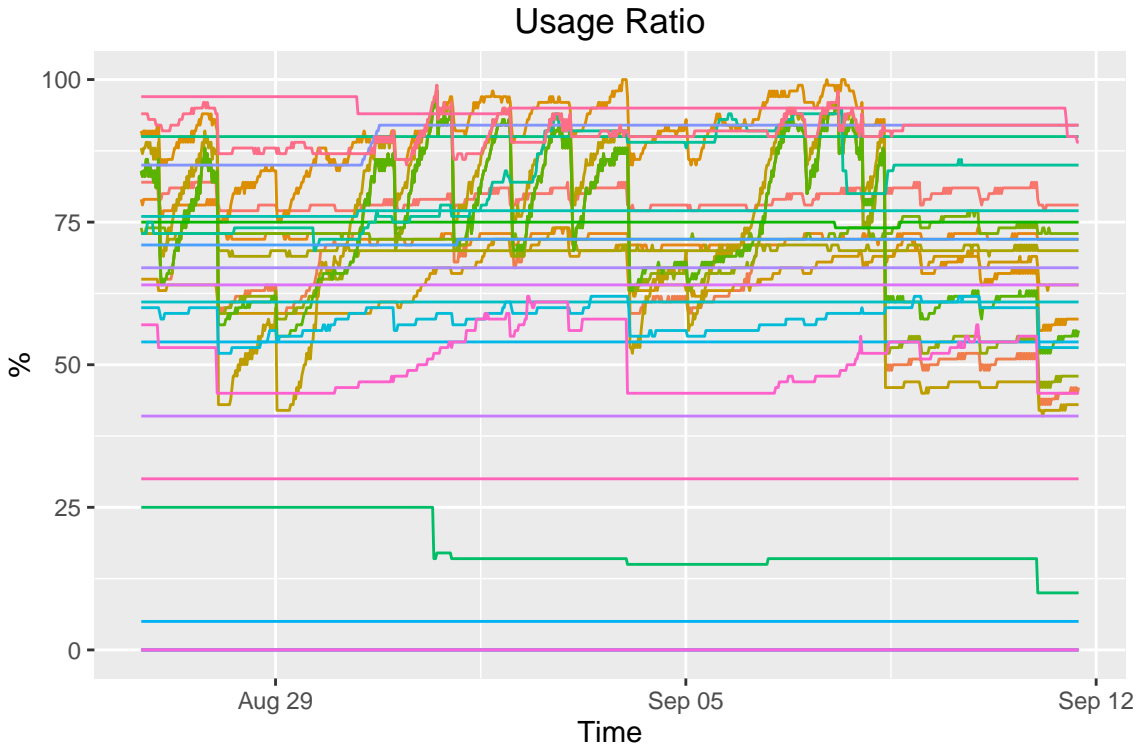
Volume



The top 5 volumes using the most capacity, and that are the most active, are the Cumulus volume, as we expected.



According the Model one volume, vs1_cumulus_clus3, has shown the biggest increase during its activity



The volumes vs1_labvdi2, vs1_ftlvmw1, vs1_xspool3, vs1_ftlvm2_vol are above the 85% threshold and that happens during the all period of the data collection.