Our cluster is on an IBM Blade Center H solution:

It has capacity for 14 nodes. We are using 12 of those 14 nodes for the Guoda cluster.
Each node (which is a vertical slice) has capacity for 2 SATA drives.
At present, each node has installed 2 x 500 GB drives, occupying both slots.
A logical partition is installed on top of both drives

Disk /dev/sda: 929.5 GiB

Operating System:
Ubuntu 5.4.0-6ubuntu1~16.04.5 (Ubuntu Xenial)
Kernel: Linux version 4.4.0-112-generic

Space Requirements

https://github.com/bio-guoda/guoda-services/issues/10
The answer to this is Yes, as soon as the cluster is online.

We are using 7 TB, ~2 of which are devoted 2018 iDigBio backups. (see https://github.com/bio-guoda/guoda-services/issues/68)

If we think that the useful data can double in these next two years => 10 TB of data
We need to leave space for temporary computations and backups => 4 TB.
14 TB in total.
But 20% is lost or dedicated to system files. => 1.2 x 14 TB = 16.8 TB

At present, we have 1 TB per machine: 2 x 512GB drives working in a single logical drive.
I believe this configuration is not optimal.

Nodes	Slot	System Board	
mesos01			
mesos02	3	68Y8161 -> HS22 (5600)	
mesos03	4	68Y8161 -> HS22 (5600)	
mesos04	5	68Y8161 -> HS22 (5600)	
mesos05	6	68Y8161 -> HS22 (5600)	
mesos06	8	68Y8161 -> HS22 (5600)	
mesos07	9	68Y8161 -> HS22 (5600)	
mesos08	10	68Y8161 -> HS22 (5600)	
mesos09	11	68Y8161 -> HS22 (5600)	
mesos10	12	68Y8075 -> HS22 (5600)	
mesos11	13	94Y8600 -> HS22 (5600)	
mesos12	14	68Y8161 -> HS22 (5600)	

There was a problem last time salt-minion was installed:

```
root@mesos12:~# apt-get install sdparm
E: dpkg was interrupted, you must manually run 'dpkg --configure -a' to correct the problem.
root@mesos12:~# dpkg --configure -a
dpkg: error processing package salt-minion (--configure):
 package is in a very bad inconsistent state; you should
 reinstall it before attempting configuration
Errors were encountered while processing:
 salt-minion
```

https://docs.saltstack.com/en/latest/

```
sudo dpkg --remove --force-remove-reinstreq salt-minion
sudo apt-get install salt-minion
```

I "solved" this way, keeping previous configuration (option N). But I do not about Salt and I have not tried if the services are correctly running.

```
root@mesos12:~# lsblk
NAME   MAJ:MIN RM   SIZE RO TYPE MOUNTPOINT
sda      8:0    0 929.5G  0 disk
├─sda1   8:1    0 905.5G  0 part /
├─sda2   8:2    0     1K  0 part
└─sda5   8:5    0    24G  0 part [SWAP]
```

```
root@mesos12:~# fdisk -l
Disk /dev/sda: 929.5 GiB, 997998985216 bytes, 1949216768 sectors
Units: sectors of 1 * 512 = 512 bytes
```

```
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: dos
Disk identifier: 0xd7ad0ea0

Device     Boot      Start        End    Sectors   Size Id Type
/dev/sda1  *          2048 1898907647 1898905600 905.5G 83 Linux
/dev/sda2       1898909694 1949214719   50305026    24G  5 Extended
/dev/sda5       1898909696 1949214719   50305024    24G 82 Linux swap / Solaris
```

```
root@mesos12:~# sdparm -a /dev/sda2
    /dev/sda2: LSILOGIC  Logical Volume    3000
```

```
root@mesos12:~# sdparm -a /dev/sda5
    /dev/sda5: LSILOGIC  Logical Volume    3000
```

Therefore, we see a logical of 905 GB created with the BladeCenter LSI Logic Volume Configuration Utility:
https://bladecenter.lenovofiles.com/help/index.jsp?topic=%2Fcom.lenovo.bladecenter.hs22.doc%2Fdw1iu_t_configuring_a_sas_raid_array.html

Cheap/Not that bad option:

- Keep one of the 512 GB drives, install:
    - A 1TB drive Part Number 81Y9730. In ebay these drives are "too" cheap ($160): https://www.ebay.com/p/IBM-1000GB-Internal-7200RPM-2-5-81Y9730-HDD/109470503
    - IBM SAS 1.2 TB 00AD075, 10K rpm.
- The 512 GB drive would run the OS and the 1TB drive would be entirely dedicated to data. (Better performance).
- This will give us 12 TB of the optimal 17 TB we want.
- Considering new collections and data sets are not added so fast, it may be an enough option
- Mesos01 and mesos02 are running as namenodes and datanodes. But we need them to do this double work.
-

Cheapest/Riskiest option:

- Do not buy anything.
- Space: After deleting 2018 data, we have 50% utilization.
    - Not much growing/temporary storage capacity (2.5 TB).
    - Data is usually not perfectly balanced
- Performance: OS and Data live in the same physical drive/controller.

Ideal:

- SSD for OS. Model 49Y6129, 200 GB, ~$1,200 each
- SSD for Data. Model 49Y6195, 1.6 TB, ~$2,000 each.

Problems in mesos11

```
root@mesos11:~# fsck -A /dev/sda1
fsck from util-linux 2.27.1
e2fsck 1.42.13 (17-May-2015)
/dev/sda1: recovering journal
ext2fs_check_desc: Corrupt group descriptor: bad block for block bitmap
fsck.ext4: Group descriptors look bad... trying backup blocks...

/dev/sda1: ***** FILE SYSTEM WAS MODIFIED *****
```

```
root@mesos11:~# dmesg | tail
[623516.421517] sd 0:1:4:0: [sda] tag#1 Add. Sense: Unrecovered read error
[623516.421523] sd 0:1:4:0: [sda] tag#1 CDB: Read(10) 28 00 69 bf bb a8 00 00 08 00
[623516.421529] blk_update_request: critical medium error, dev sda, sector 1774173096
[623516.421585] Buffer I/O error on dev sda1, logical block 221771381, async page read
[623516.421730] systemd-journald[315]: Failed to write entry (12 items, 367 bytes), ignoring: Read-only file system
[623516.421882] systemd-journald[315]: Failed to write entry (12 items, 348 bytes), ignoring: Read-only file system
[623516.422032] systemd-journald[315]: Failed to write entry (12 items, 347 bytes), ignoring: Read-only file system
[623516.422181] systemd-journald[315]: Failed to write entry (12 items, 356 bytes), ignoring: Read-only file system
[623516.422254] systemd-journald[315]: Failed to write entry (9 items, 287 bytes), ignoring: Read-only file system
[623516.422330] systemd-journald[315]: Failed to write entry (9 items, 288 bytes), ignoring: Read-only file system
```