Configuration 1: Hadoop/Spark Cluster with 128K-cores, 1PB memory, 200PB HDD, and 10Gb/s Ethernet Fat-Tree network (each VM should be equivalent to the d2.8xlarge instance); in addition to the compute resources, a 400PB distributed storage shared across the entire cloud should be procured, with enough capacity for 400GB/sec throughput (for pricing comparison, see S3)

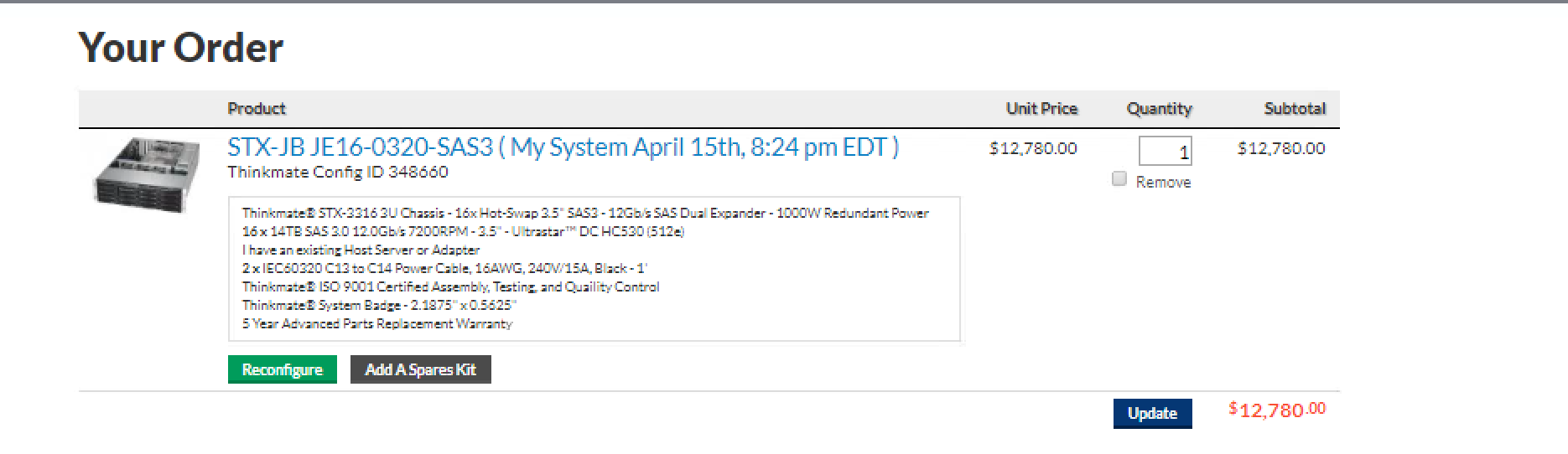


Fig 1: -JBOD

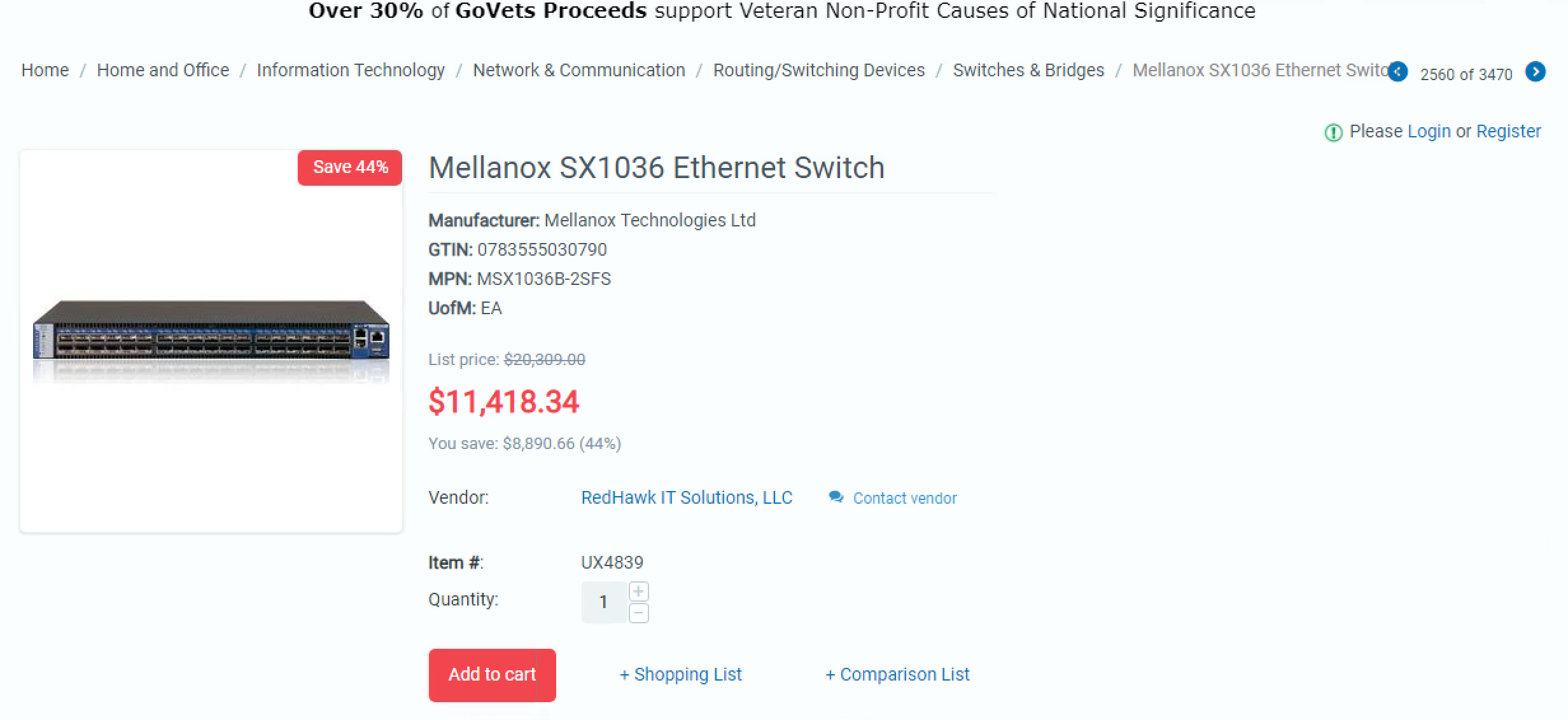


FiG-2:Switch

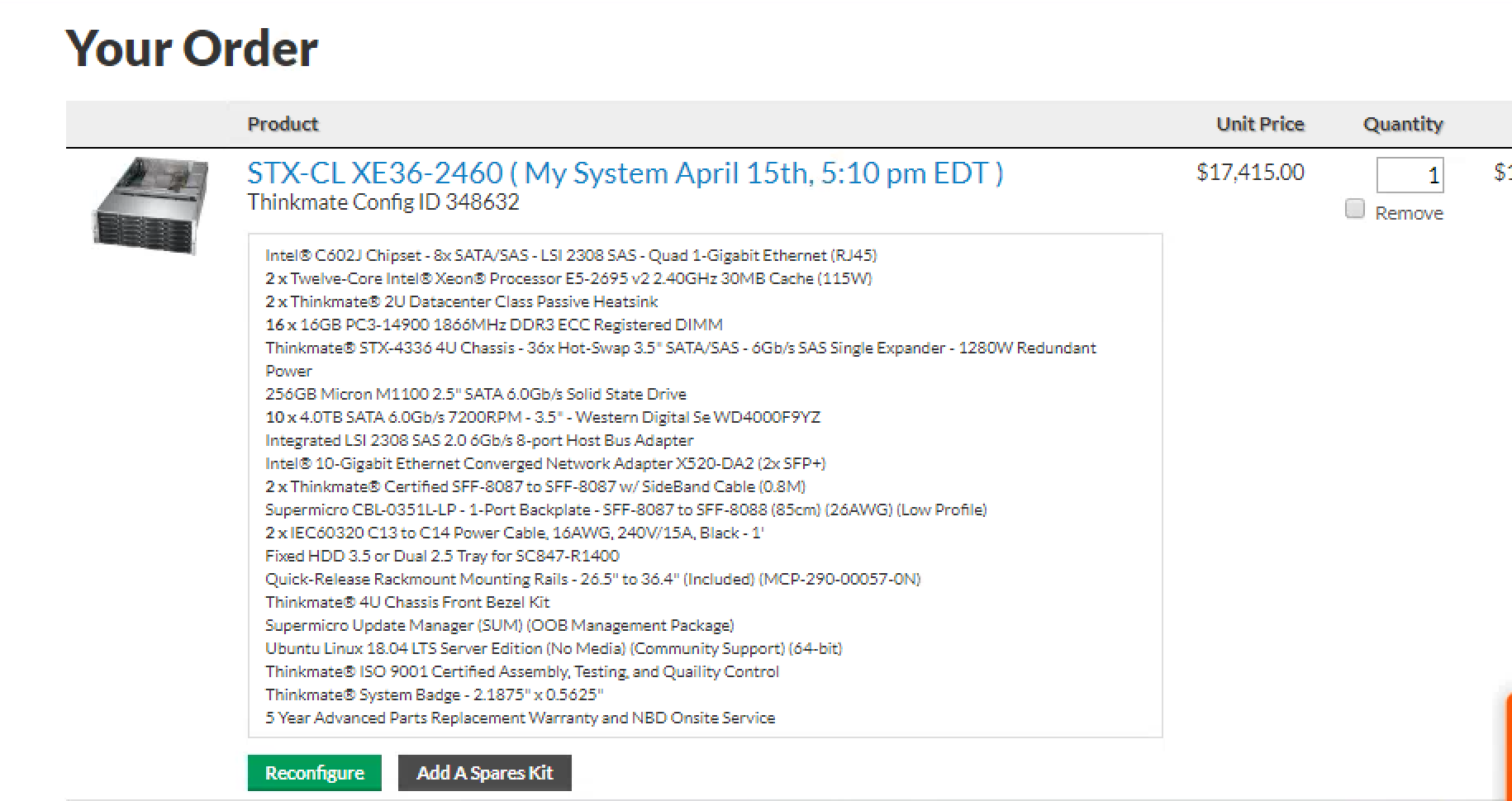


FIG-3 servers.

128K cores means 3556number of d2.8xlarge EC2 instance.

5.52 dollar / hour pricing.

HW7

====

we have to create 1 PB of memory and 200PB of harddisk and 10Gb/s fattree network. and 400PB of distributed storage with 400 GB/sec

throughput.

1PB mem 200 PB hardsik 10Gb/s fatree and 400PB of distributed storage.

d2.8xlarge instance config

==========================

cores ECU mem hard disk price (Linux on demand)

d2.8xlarge 36 116 244 GiB 24 x 2000 HDD $5.52 per Hour

5.52 is on demand price .

1PB = 1000TB = 1000 \* 1000 GB = 10^6 GB

200 PB = 200 \* 10^ 6GB of hard disk.

d2.8xlarge has 48000 GB of harddisk (This is in decimal value) = 48 TB.

1PB = 1000TB / 48TB = 21

200PB = 21 \* 200 = 4200 servers

but as per cores this should be 3556 number of servers.

calculation: 128K cores, d2.8xlarge has 36 vCPU, so 36 cores.

128000/36 = 3556 servers or d2.8xlarge EC2 instances.

now we have 1PB of RAM , so that is 1PB = 1000 TB = 10^6 GB / 244 GiB = 4096 servers.

highest value is 4200 to serve the need of 200PB of HDD, so we have to pay for the access capacity that we are going to purchase.

so to keep the configuration consistent.

we also need 400PB of distributed storage which we attach to servers as S3 storage assuming that I/O would be frequent as the servers

are used for hadoop/spark so, we can use s3 standard storage type.

400 PB of which will cost as per the below table.

S3 Standard Storage Pricing

First 50 TB/mo $0.023 per GB

Next 450 TB/mo $0.022 per GB

Over 500 TB/mo $0.021 per GB

we have total 4200 servers (d2.8xlarge EC2 instances): 400PB = 400 \* 1000 TB = 400 \* 1000 \* 1000 GB ,

so per instance , 400 \* 1000 \* 1000 = 95326 GB = 95.3 TB we can take this as 100 TB per instance for easier calculation ,

also the price remain same, so, 100 TB = 100 \* 1000 GB \* 0.022 $ for one server, so for 4200 servers

= 100\* 1000GB \* 0.022 \* 4200 $

= $9240000 /per month

Total server 4200 \* 5.52 = $23184 /per hour.

so in 5 years = $23184 \* 5\*365\*24 = $1015459200

TCO : $1015459200 + $9240000\*5 = $10246992000

amazon s3 standard throughput in GB/sec is not definitive , but it is able to do 3500 PUT/POST/DELETE

**Request Rate and Performance Guidelines**

Amazon S3 automatically scales to high request rates. For example, your application can achieve at least 3,500 PUT/POST/DELETE and 5,500 GET requests per second per prefix in a bucket.

<https://docs.aws.amazon.com/AmazonS3/latest/dev/request-rate-perf-considerations.html>

if we take 32KB size of each PUT/POST/DELETE that would be 3500 \* 32KB = 112000KB = 112 MB /second

considering there are 4200 servers so 4200 \* 112MB = 470400 MB = 470 GB which more than what we need for cluster wide throughput.

we can not create a FAT-TREE network but we can host this in single availability zone , so to maximise network speed.

FROM think-mate (private cloud) machine model (STX-CLXE36-2460)

========================

each node with 2\*12 core number of processors = 24 core. and storage 24 \* 4TB = 96 TB (this is 2 \* d2.8xlarge)

Reason for going with kind config : storage heavy machine tends to cost less then Hight density machine that is machine with high socket and processor count.

so the overall cost will come down even though number of machines would be more.

so , 128000 core / 24 core = 5333 number of machine

with storage 40 TB per node. so 5333 \* 40 = 213320 ~ 200PB

with 10GBE ethernet with QSFP

RAM : 16\*32 = 256 GB (should be 187 GB per node to fulfil 1PB requirement but config is not available in thinkmate)

ToTal cost per server = $17415 . This is including 5 year parts replacement warranty.

$17415 \* 5333 = $92874195

Total wattage reported 516.8 watt.

SWITCH COST

===========

*Example. Connect N=128 nodes, with Pe=Pc=36 ports and a blocking factor Bl=1. Calculate E and C, the number of edge and core switches, respectively.*

*As Bl=1, Eptn=Eptc=Pe/2=18.*

*Then, E=Ceil (N/Eptn)=Ceil (7,1)=8.*

*B=Pc div E=36 div 8=4.*

*C=Ceil (Eptc/B)=Ceil (18/4)=Ceil (4,5)=5.*

*The answer is: E=8, C=5.*

*replace N = 128 to 5333 , so , E = Ceil(5333/18) = 297*

*B = 36 / 297 = 0.12*

*C = Ceil(18/0.12) = 150*

*so core switches = 150 and edge switches = 297.*

*total number of switch = 297 + 150 = 447 switches.*

*SX1036 cost = $11418 \* 447 = $5103846*

*power consumption is 86 WATT for mellanox switches.*

*Cable cost*

*========*

*if we take 36 QSFP to QSFP cable per node , then total cable = 447 \*36 = 16092 number of cables.*

[*http://www.cablesondemand.com/category/QSFP%2B/URvars/Catalog/Library/InfoManage/QSFP+\_CABLES\_(QSFP\_PLUS).htm*](http://www.cablesondemand.com/category/QSFP%2B/URvars/Catalog/Library/InfoManage/QSFP+_CABLES_(QSFP_PLUS).htm)

*according to the above website QTY > 50 = $71.75*

*so 16092 \* $71.75 = $1154601 dollar.*

*Taking Rack size as 42U and price which is $875 per RACK from this website :*[*https://www.fs.com/products/73958.html?currency=USD&paid=google\_shopping&gclid=Cj0KCQjw19DlBRCSARIsAOnfReibBMMwz\_IDlg\_djaqCkx6phHzC\_Or\_zeE-ThutV2\_J0WfR00d3QusaAo8NEALw\_wcB*](https://www.fs.com/products/73958.html?currency=USD&paid=google_shopping&gclid=Cj0KCQjw19DlBRCSARIsAOnfReibBMMwz_IDlg_djaqCkx6phHzC_Or_zeE-ThutV2_J0WfR00d3QusaAo8NEALw_wcB)

*200 PB extra storage(JBOD)*

*================*

*have to configure NAS storage on this , so we can use JBOD.*

*if we choose the system with 224 TB of storage we have to buy 893 JBOD. ~ 200 PB .*

*$12780 per JBOD \* 893 = $11412540*

*RACK Cost*

*========*

*our machine :* STX-CLXE36-2460 is a 6U machine . and switch SX1036 is a 1U machine.

total U = 5333 \* 6U + 447 \* 1U = 31998 + 447 = 32445 / 42 = 772.5 ~ 773 number of racks.

$875 \* 773 = $676375

power cost

========

516.8 WATT = 516.8/1000 = 0.516 KWH

10 Cents per KHW ,

so 10 \* 0.516 = 5.1 cents per hour \* 5333 servers = 27198.3 per hour = 272 dollar per hour.

272 \* 5 \* 365 \* 24 = $11913600 in 5 years.

86 WATT for the switches 86 /1000 = .086 KWH \* 10 cents = .86 cents per hour /per switch.

.86 \* 447 = 385 cents or 3.85 dollar per hour

3.85 \* 5 \* 24 \* 365 = $168630 in 5 years

Total power cost = $11913600 + $168630 = $12082230 in 5 years ,

we have to take same amount in cooling as well , so $12082230 \* 2 = $24164460

administrator cost

==============

$100000/Year , total server 5333 server + 447 switch = 5780 /1000 = 5.78 ~ 6 administrator = $600000/per year \* 5 = $3000000

TCO in 5 years

===========

$24164460—> power + cooling

$676375 ——> racks

*$1154601 ——> cable*

*$5103846——> switch*

$92874195—>server cost

$3000000—> administrator cost

*$11412540—> JBOD cost*

================

TCO: $138386017

Explain in words if it is better to rent or buy. If it is better to buy, what utilization must you maintain over the 5 year lifetime of the private cloud in order to break even on the investment?

**Creating private cloud is two order of magnitude cheaper than renting from AWS.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Description | Price per Item | Quantity | Total Price |
| Compute Servers | STL-CL XC36 | $17415 | 5333 | 92874195 |
| Network Switches | SX1036 | $11000 | 447 | *5103846* |
| Network Cables | QSFP | $77 | 16092 | *1154601* |
| Racks | 42U GR600-Series Black Server Cabinet 600x1170mm with 2 PDU Brackets and Adjustable Fixed Shelves | $875 | 773 | 676375 |
| Storage Servers | STX-JB | $12780 | *893* | *$11412540* |
| Electric Power |  |  |  | $12082230 |
| Cooling |  |  |  | $12082230 |
| Administration |  |  |  | $3000000 |
| TOTAL | N/A | N/A | N/A |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Configuration 1 | Configuration 2 | Configuration 3 |
| Public Cloud (including EC2 and S3) Cost over 5 years, 24/7 operation, with 100% usage | $10246992000 |  |  |
| Private Cloud cost over 5 years, 24/7 operation, with 100% usage  What utilization must be achieved with the private cloud to make the private cloud option more attractive than the public cloud? | Private cloud is better than renting any way. |  |  |
|  |  |  |  |