Re: Disck write latency > 50ms

Source:
http://www.tech−archive.net/Archive/Exchange/microsoft.public.exchange.admin/2006−02/msg03313.html

- *From*: "John Fullbright" <fullbrij@xxxxxxxxxxx>
- *Date*: Thu, 23 Feb 2006 10:13:12 −0800

Exmon.

"SW" <SW@xxxxxxxxxxxxxxxxxxxxxxxxx> wrote in message
news:49A5BD65−D062−45D9−B922−7D328F6FF229@xxxxxxxxxxxxxxx

> Thanks, is there any possible way to find out what user causes the
> "slowness"? would be nice to see some thing like domain\username?
>
> Also how can I tell how many users are connected to the Exchange 2003
> server?
>
> Thanks
>
> "John Fullbright" wrote:
>
> > I use the figure 85 IOPS (IO operations per second) for a 10K spindle.
> > Some
> > folks use the value 130, which is derived by tahing the average seek time
> > plus the rotational latency, and dividing 1 second by the sum. This does
> > not account for the response time. Yes, a 10K spindle can handle more
> > than
> > 130 IOPS, however at 130 IOPS, the response time is roughly 60 − 80ms.
> > 85
> > IOPS is at a target 20ms IO.
> >
> > Physical Disk − avg sec/write, sec/read, and sec/transaction tell you how
> > long IO operations against a given physical disk are taking. MOM uses
> > the
> > criteria from "Optimizing Storage for Exchange Server 2003" to alert;
> > "average 20ms writes with no peaks lasting more than a few seconds over
> > 50ms".
> >
> > Physical Disk − Transactions/sec, read/sec, write/sec tell you how many
> > IOs

and of what type a physical disk is servicing in a second.
Transactions/sec
translates to IOPS. You can use these three counters to calculate your
actual read/write ratio.

Database − Log Record Stalls/sec. This tells you haow many times the log
buffers filled to the high water mark, and a forced commit commenced.
During a forced commit, all cliet IO is quiesced. This value should not
exceed an average of 10 or peak of 100. Exceeding this value generally
means your log disk is inadequate.

Database − Page fault stalls/sec. This tells you how many times per
second
new pages were not able to be read into deatabase buffers in memory.
Generally this occurs when all pages are locked; either assembling
transactions or awaiting transfer to the log buffers. Any value greater
than zero is a sign of a serious IO bottleneck.

"SW" <SW@xxxxxxxxxxxxxxxxxxxxxxxxx> wrote in message
news:C4AD0940−C4CC−4105−96AE−C61C066D3C61@xxxxxxxxxxxxxxx

> At the moment I am only monitoring Physical Disk Avg.
> Disk sec/write
> and
> seing if it goes over 0.02 (20ms) and 0.05 (50ms). What esle
> should I
> check
> and what should the warning thresholds be?
>
> "Vicnice" wrote:
>
>> Input and Output Operations per Second.
>> Basically, anytime that
>> Exchange
>> has
>> to read or write that is an IOP. The standard
>> is 2:1 (two reads one
>> write −
>> people are reading more email than they are
>> usually sending).
>>
>> With that being said an average IOP that a
>> single disk can handle is
>> between
>> 80−140 per second. That means that it can
>> do 80 to 140 operations per
>> second.
>> When users exceed this, it is good bye good

performance and HELLO
popup
city
(Outlook is waiting for Communication with
the Exchange Server).

http://www.microsoft.com/technet/prodtechnol/exchange/guides/StoragePerformance/

There you go.

Honestly the calculations made no sense to
me as I am a bit
"inexperienced"
(read:stupid) compared to the other people
here. I would like an
explaination
to the formula as well.

"SW" wrote:

> John that is quite amazing.
> We will go to RAID 10, we
> have enough
> bays
> to
> put in extra disks, we are to
> but another RAID card for
> this too.
> You
> calculations are no doubt
> correct, can you explain
> how you got those
> results,
> I just want to explain to my
> bosses. What does IOPS
> stand for?
>
> "John Fullbright" wrote:
>
> > Dump the
> > RAID 5 and
> > go RAID
> > 10. The
> > problem is
> > write
> > performance
> > on

Drive
F:

Let's
assume a 4
disk RAID
5 vs. a 4
disk RAID
10 array.
Further,
let's
assume 10K
RPM SCSI
drives and a
2:1
read/write
ratio.

RAID 5:

Write perf =
P*N'/4 =
85*3/4 =
63.75 IOPS
read perf =
P*N' =
85*3 = 255
IOPS
mixed perf
=
255*.66+63.75*.33
= 168.3 +
21 = 189.3
IOPS

RAID 10

write perf =
P*N/2 =
85*4/2 =
170 IOPS
read perf =
P*N = 85*4
= 340 IOPS
mixed perf
= 340*.66 +
170*.33 =
224.4 +
56.1 =
280.5 IOPS

That's a 100% increase in write performance and a 55.5% increase in overall performance for RAID 10 for the same number of spindles. I would never recommend RAID 5 for exchange.

"SW" <SW@xxxxxxxxxxxxxxxxxxxxxxxxx> wrote in message news:E072839C−D358−40B3−B67D−CA89B83840A2@xxxxxxxxx

It's the "writes" that are the problem, and when this is high our users "nag". I use MOM 2005

with
the
Exchange
Management
Pack
which
alerts
me
when
the
Physical
Disck
Avg.
Disk
sec/write
is
greater
than
50ms
for
10
minutes
(it
polls
every
60
seconds).
Here
is
one
I
got
earlier:

Severity:
Warning
Status:
New
Source:
PhysicalDisk:
Avg.
Disk
sec/Write:
0
F:
Name:
Disk
Write
Latencies
>
50

msec
(F:
Drive)
Description:
High
Disk
Write
Latencies
for
the
past
10
minutes
PhysicalDisk:
Avg.
Disk
sec/Write:
0
F:
value
=
6.41240062645155E−02.
The
average
over
last
10
samples
is
0.064124.
Domain:
DOMAIN
Agent:
EXCHANGE2003
Time:
2/22/2006
09:57:00

"Neil
Hobson
[MVP]"
wrote:

> RAID5
> probably
> isn't
> the
> best
> answer.

Are you using MOM to monitor this server? Is it actually causing a problem to the users? In a nutshell, you should look to size your server correctly regarding disk IOPS. There are a number of articles on doing this, such as those by Rui Silva at msexchange.org, or Microsoft's own

paper
called
something
like
Optimizing
Storage
for
Exchange
2003.

−−
Neil
Hobson
Exchange
MVP

"SW"
<SW@xxxxxxxxxxxxxxxxxxxxxxxxx>
wrote
in
message
news:582F4999−F46F−4DAA−9B59−E60B3227B36

> What
> can
> we
> do
> about
> disk
> latency
> that
> goes
> over
> 5ms
> at
> various
> time
> of
> the
> day?
>
> We
> have
> an
> Exchange
> 2030
> Ent
> server
> (400
> users),

2
DB's
80gb
and
10GB.

The
high
latency
is
on
our
F
drive,
he
is
our
setup:

c:
OS,
Exchange
program,
paging
file
(RAID
1
channel
0)
e:
Transaction
logs
(RAID
1
channel
1)
f:
Databases
(RAID
5
on
a
separate
RAID
card)

Usually
the
server
runs
below

20ms
on
the
Avg.
Disk
sec/write
what
we
hit

50ms
at
stages
throughout
the
day,
especially
first
thing
when
everyone
logs

in.

.