

Re: FTS Performance in SQL 2005

Source:

<http://www.tech-archive.net/Archive/SQL-Server/microsoft.public.sqlserver.fulltext/2006-11/msg00014.html>

- *From:* "Hilary Cotter" <hilary.cotter@xxxxxxxx>
 - *Date:* Fri, 3 Nov 2006 06:36:58 -0500
-

can you do this for me and post the results back here

```
sp_configure 'max server memory (MB)'
```

I don't think you have left enough memory for the OS and MSSearch.

Hilary Cotter
Director of Text Mining and Database Strategy
RelevantNOISE.Com – Dedicated to mining blogs for business intelligence.

This posting is my own and doesn't necessarily represent RelevantNoise's positions, strategies or opinions.

Looking for a SQL Server replication book?
<http://www.nwsu.com/0974973602.html>

Looking for a FAQ on Indexing Services/SQL FTS
<http://www.indexserverfaq.com>

<tony.newsgrps@xxxxxxxx> wrote in message
news:1162417354.892553.92200@xx

Thank you for your answer. We'll look at partitioning our table if we get into similar problems.

One quick follow up:

We keep on growing our table and the performance keeps on dropping. With 8 million rows we had about 15 queries/sec. With 20 million rows, we dropped to about 2 queries per sec. The server seems completely underused though. The CPU and memory usage are very low and we see a lot of page faults. Any idea what happened and how we could get back to 15 queries/sec? Could it be that the index needs to be re-organized or something like that?

Our table is very simple (2 fields: 1 id, 1 plain text) and our queries match only a very limited set of documents (100 matching records max

Re: FTS Performance in SQL 2005

out of 20 millions).

Also, with your 300 millions/2TB table, what performance do you get on queries (on average) and on what hardware?

Thanks a lot for your guidance.

Tony.

Hilary Cotter wrote:

Basically we find that the queries start taking longer and a bounce seems to improve performance.

I have not communicated this to MS. You might want to open a support incident yourself.

When we had a single table of over 300 million rows and pushing 2 terabytes we had no end of problems with SQL FTS. After breaking the table up into 50 million row partitions we have had no real problems, but still bounce fts weekly.

--

Hilary Cotter
Director of Text Mining and Database Strategy
RelevantNOISE.Com – Dedicated to mining blogs for business intelligence.

This posting is my own and doesn't necessarily represent RelevantNoise's positions, strategies or opinions.

Looking for a SQL Server replication book?

<http://www.nwsu.com/0974973602.html>

Looking for a FAQ on Indexing Services/SQL FTS

<http://www.indexserverfaq.com>

<tony.newsgrps@xxxxxxxxxx> wrote in message
<news:1162399142.153739.316510@xx>

Hi Hilary,

Your comment here is a bit scary. It sounds like the FTS capabilities of Sql Server 2005 are not ready for production. Can you detail a bit more the problems you encounter that force you to restart the sql fts

Re: FTS Performance in SQL 2005

once a week? Is MS aware of that problem? What are their recommendations? Do you know of any upcoming patch or SP that would fix this?

Tony.

Hilary Cotter wrote:

We pound full-text search the same way you do. There are advantages to a multi-proc machine – a quad or eight way. We have to restart sql fts once a week. We find that smaller tables work better – where smaller is 50 million or so rows.

We also found the following settings work well:

setting a high resource usage to 5 and reorganize frequently.
set ft crawl bandwidth (max) and ft notify bandwidth (max) to 0,
set max full-text crawl range to the number of cpu's on your system,
index text only,
put your catalogs on the fastest disk subsystem (RAID 10) possible preferably with their own controller,
and run 64 bit.

--
Hilary Cotter
Director of Text Mining and Database Strategy
RelevantNOISE.Com – Dedicated to mining blogs for business intelligence.

This posting is my own and doesn't necessarily represent RelevantNoise's

Re: FTS Performance in SQL 2005

positions, strategies or opinions.

Looking for a SQL Server replication book?

<http://www.nwsu.com/0974973602.html>

Looking for a FAQ on Indexing

Services/SQL FTS

<http://www.indexserverfaq.com>

"Doug Funk"

<doug.funk@xxxxxxxxxxxxxxxxxxxxxx>

wrote in message

[news:IAN_g.60102\\$OI1.44332@xxxxxxxxxxxxxxxxxx](mailto:news:IAN_g.60102$OI1.44332@xxxxxxxxxxxxxxxxxx)

I do not see any resolution to this problem mentioned, and have a similar problem.

I have just implemented a database with 12 tables, one FT index on a text column for each table. The unique key column is a uniqueidentifier.

This is under SQL Server 2005, SP1, Windows 2003 Server, on an x64 dual processor system with 16 GB RAM. SQL Server is limited to 12 GB RAM and nothing else runs on the box.

The query uses CONTAINSTABLE.

We have an automated process that feeds thousands of queries, one at a time, to run FT searches.

It appears that the searches

Re: FTS Performance in SQL 2005

run fine for a while, then the searches take longer and longer, until eventually the search never returns, for hours anyway.

When we see this happening, in Windows Task Manager we see that process msftesql shows PF Delta up over 50,000.

The Memory Usage and VM Size never increase over about 65 MB and 20MB.

Did you ever find a solution for this ?

Thanks.

Doug Funk
News Data Services
dfunk@xxxxxxxxxxxxxxxxxxxxx

"Simon Sabin"
<SimonSabin@xxxxxxxxxxxxxxxxxx>
wrote in message
news:c4366deffa728c87fe6f490db50@xxxxxxxxxxxxxxxxxxxxxxxxxx

Hello
KaMa,

The maximum equates to process ~4.5GB/s thats a lot.

Can you post you query plans and the output of

Re: FTS Performance in SQL 2005

statistics IO
Simon
Sabin
SQL Server
MVP
<http://sqlblogcasts.com/blogs/simons>

Hello
Simon,
The
PageLookups/sec
are
avg.
15000–17000
when
doing
a
large
full-text
query,
but
currently
I
also
have
other
processes
quering
the
FT,
so
this
could.
Minimum:
0
Maximum:
560613.508
Average:
53890.00
Values
from
~
100
seconds...

Do
you
know
if

Re: FTS Performance in SQL 2005

the
memory
FT/SQL
uses
was
merged?
Because
i've
set
the
MAX
sql-server
memory
to
3.5
GB
instead
of
4.0
GB
Simon
Sabin
schrieb:

Hello
KaMa,

its
the
pipe
between
the
CPU
and
Memory
that
could
be
the
bottleneck.
You
try
drinking
a
gallon
of
water
with
a
straw,

Re: FTS Performance in SQL 2005

the
bottleneck
is
the
straw
it
can
only
handle
a
certain
flow
of
liquid,
its
the
pipe
thats
the
bottleneck
not
you
(as
the
CPU).
Now
try
with
a
1
inch
pipe
the
bottleneck
is
now
your
ability
to
drink
all
the
fluid
that
you
can
suck,
the
pipe
is
now

Re: FTS Performance in SQL 2005

not
the
bottleneck
instead
you
are.
Have
a
look
at
the
Pagelookups/sec
in
the
Buffer
Manager.

I'm
not
saying
this
is
the
bottleneck
but
it
is
likely.

Simon
Sabin
SQL
Server
MVP
<http://sqlblogcasts.com/blogs/simons>

Hi
Simon,
But
why
is
it
that
I
(almost)
don't
see
any
change
in
Memory

Re: FTS Performance in SQL 2005

or
CPU
usage?
I
mean
if
the
software
is
running
into
a
bottleneck,
I
should
see
it
–
right?
Simon
Sabin
schrieb:

Hello
KaMa,

On
the
memory
issue,
imagine
this.
You
need
to
read
10
books
and
your
in
a
library.
The
first
time
you
want
the
book
you

have
to
go
find
it
off
the
shelves.
The
second
time
you
have
the
book
on
your
desk.
Now
imagine
the
process
of
reading
the
data,
you
can
only
read
the
books
at
a
certain
speed.
Even
though
you
don't
have
to
go
hunting
for
the
book
on
the
shelves
you

still
need
to
read
each
page,
if
the
book
is
big
it
is
going
to
take
longer.
In
this
situation
the
library
shelves
are
the
disks,
your
desk
is
the
memory,
and
you
are
the
CPU.
You
have
a
finite
limit
for
reading
pages
this
is
the
memory
bandwidth
between
the

CPU
and
the
memory.
With
Full
text
their
is
an
additional
bit
and
that's
the
communication
with
full
text
engine,
you
can
almost
think
of
this
as
someone
else
reading
the
books
to
and
writing
the
page
numbers
you
want
on
a
bit
of
paper
and
then
you
having
to
read

those
pages.
So
simply
put
you
have
to
reduce
the
amount
of
data
you
read
to
get
performance
Simon
Sabin
SQL
Server
MVP
<http://sqlblogcasts.com/blogs/simons>

Full-Text
Table:
id
-
int
-
primary
key
-
clustered
(FULL
TEXT
KEY)
srcID
-
varchar(30)
(LINKING
KEY
TO
ARTICLE)
srchField
-
varchar(150)
-
FULL-TEXT
INDEXED

```
src
-
varchar(15)
(SOURCE
OF
KEYWORD)
updateStamp
-
timestamp
(LOCAL
TIMESTAMP)
Indexes
(fields
/
description)
0
-
[id]
-
Primary
Key
-
unique
/
clustered
1
-
[id],
[srcID]
-
unique
/
nonclustered
2
-
[srchField]
-
non-unique
/
nonclustered
3
-
[srcID]
-
unique
/
nonclustered
4
-
[id]
-
```

non-unique
/
nonclustered
Index
in
use
when
using
a
CONTAINS
or
CONTAINSTABLE
(1).
I
can't
imagine
that
the
memory
is
a
bottleneck,
because
I
do
not
see
any
change
in
memory
usage
when
I
use
a
fulltext
query...
Do
you
think
i
should
reduce
the
max
memory
of
sql
server
to

<3.5
GB
(currently
set
to
3.5
GB
of
4
GB)?
Another
question
–
what
means
DDL?
:)
Thanks!
Simon
Sabin
schrieb:

The
article
title
is
won't
help
either
as
the
full
text
engine
doesn't
return
this.

The
reason
that
the
CPU
usage
is
like
that
is
probably
because

you
bottleneck
is
memory.

FTS
returns
all
the
keys
that
match
the
criteria
it
will
then
join
to
your
main
table
for
each
key
returned,
if
the
indexing
is
not
right
then
you
will
end
up
reading
all
the
data
from
the
main
table.
Whilst
this
may
be
cached
it

still
has
to
be
processed,
if
its
not
cached
it
will
be
being
read
from
disk
which
is
worse.

Your
only
option
is
to
ensure
that
you
are
only
returning
the
records
you
need
from
the
FT
engine.
i.e
all
criteria
is
passed.
and
that
you
have
a
covering
index

on
the
columns
in
the
query.

Can
you
please
post
the
DDL
for
the
table
and
the
indexes
on
the
table.

Simon
Sabin
SQL
Server
MVP
<http://sqlblogcasts.com>

Hello
Dan,
I
know
that
the
upper
queries
are
mostly
extremes,
but
I
don't
want
do
construct
a
code
wich
can

run
into
timeouts
or
searches
which
take
~
25–40
seconds
–
even
in
the
hardest
cases...
The
problem
with
partitioning
is
that
I
need
to
give
different
ways
of
ordering
top
returned
articles
(e.g.
views,
date..) and
so
i
cannot
do
several
full-text
catalogs
wich
access
different
tables...
The
problem
with

the
ranking-order
is
that
none
of
my
order-clause
use
the
searched
fields
itself
(except
ordering
for
the
article
title),
and
i've
tried
CONTAINS
with
TOP_N_RA
with
some
tests
and
got
completely
wrong
results...
No
partial
search
-
here
the
problem
is
that
I
have
some
words
which
occur
several
thousand
times,

the
fulltexttable
is
constructed
like
a
keyword
list
–
so
if
someone
searched
for
a
keyword
that
occurs
several
times
–
the
problem
would
occur
again...
I've
already
implemented
a
caching
scenario
to
the
whole
thing
which
caches
statistics,
counts
and
results
of
the
search...

Do
you
by
any
chance

know
how
to
give
the
SQL-FT
service
more
priority?
Because
what
really
buggers
me
is
that
if
I
execute
a
big
query
on
FTS,
almost
no
CPU/Memory
is
used
by
the
service
at
all
-
I
didn't
find
any
documentati
on
the
memory
behaviour
of
SQL2005/FT
compared
to
SQL2000/FT
-
because

in
SQL2000/FT
you
had
to
fix
the
maximum
of
ram
SQL
uses
to
give
FT
more
memory...

I've
already
tried
to
create
a
multiple-col
scenario,
but
this
didn't
work
out
due
to
following
problem...

```
id  
|  
srcID  
|  
title  
|  
alternate_titl  
|  
keywords  
|  
names
```

Now
if
i've

a
field
'names'
containing
–
let's
say
–
following
data:
KaMa,
Daniel
Chrichton,
Luis
Mortan

Now
if
I
search
for
'Daniel
Mortan'
–
the
record
would
appear
even
if
the
data
is
not
correct...
This
way
i
could
reduce
the
number
of
records
returned,
but
still
the
problem
persists
that

this
wouldn't
be
scalable,
because
if
some
keywords
would
appear
in
a
lot
of
records
–
there
still
are
a
lot
of
records
to
return
:).
Daniel
Crichton
schrieb:

KaM
wrot
on
18
Jul
2006
01:4
–070

Re: FTS Performance in SQL 2005

See
stran
if
it
is,
as
you'
doin
prett
muc
wha
a
strai
CON
quer
and
retri
srcII
shou
be
doin

My
table
is
only
510,
rows
so
it's
not
as
large
as
your

When
retrie
just
the
[KE
that'
wha
you'
doin
I
don'
know
of
a
way
to
acce
the
FTS
serv
outs
of
SQL
Serv
if
there
is
a
way
at
all.

To
com
the
quer
put
them
all
into
Que
Ana
and
turn
on
Sho
Exec
Plan
and
then
run
the
who
lot
-
the
Exec
Plan
wind
will
then
give
you
a
perc
quer
cost
relat
to

the
who
batch
mak
it
easi
to
spot
whic
appe
to
be
the
best
solu
Basi
it
on
time
is
a
bad
idea
as
cach
can
affe
the
resu

If
you
can
do
so,
try
to
buil
your
quer
to
matc
on
who
wor
or
redu
the
num
of
resu

you
pull
back
The
more
resu
you
have
the
wors
your
perfo
will
be
as
SQL
Serv
will
have
to
look
up
addi
data
and
with
a
very
large
num
of
rows
you
end
up
with
a
Mer
Join
Join
whic
adds
over
on
the
Sort
For
exan
if
I
run

the
follo
the
quer
cost
for
each
is:

```
select  
ID  
from  
STK  
where  
CON  
"win  
Resu  
1272  
Cost  
14.8  
(nes  
loop
```

```
select  
ID  
from  
STK  
where  
CON  
"win  
Resu  
1450  
Cost  
14.9  
(nes  
loop
```

```
select  
ID  
from  
STK  
where  
CON  
"win  
Resu  
7265  
Cost  
15.2  
(mer  
join  
with
```

sort)

```
select  
ID  
from  
STK  
where  
CON  
"wi"  
Resu  
6032  
Cost  
23.7  
(mer  
join  
with  
sort)
```

```
select  
ID  
from  
STK  
where  
CON  
"w*"  
Resu  
1765  
Cost  
31.1  
(has  
matc
```

As
you
can
see,
the
more
rows
that
are
return
the
more
work
is
done
to
pull
infor
from

the
table

Com
on
the
FTS
search
itself.
it's
obvi
that
return
a
lot
of
rows
is
a
bad
idea

selec
[KE
from
CON
"wi
Cost
3.09

selec
[KE
from
CON
"wi
Cost
3.12

selec
[KE
from
CON
"wi
Cost
4.15

selec
[KE
from
CON
"wi

Cost
30.0

select
[KE
from
CON
"w*
Cost
59.6

Noti
how
the
last
two
cost
sign
more
due
to
the
num
of
rows
bein
retur
Retr
data
from
the
FTS
engi
is
a
com
slow
proc
This
is
muc
more
obvi
whe
you
use
TOP
to
redu
the
num

of
resu

```
select  
[KE  
from  
CON  
"w*  
Cost  
97.9
```

```
select  
[KE  
from  
CON  
"w*  
Cost  
2.05
```

If
you
can
find
a
way
to
work
with
the
TOP
and
CON
you
shou
be
able
to
optim
your
quer
sign
Luck
I
rare
have
quer
that
take
long
than
2

or
3
seco
to
run
due
to
only
havi
510,
rows
and
in
mos
case
less
than
2,00
resu
for
any
give
sear
I
notic
in
your
origi
post
that
you
say
that
you
can
do
this
due
to
the
resu
bein
"ord
total
wron
Ther
is
noth
stop
you
usin

an
OR
BY
on
the
return
result
to
chan
the
sort
orde
and
usin
the
ISA
NEA
and
othe
keyv
in
FTS
to
adju
the
way
the
rank
is
calcu
to
try
to
ensu
that
you
get
just
the
resu
that
you
need
from
the
FTS
sear

Dan

