

Re: Generating one table with a terabyte of data

Source:

<http://www.tech-archive.net/Archive/SQL-Server/microsoft.public.sqlserver.dts/2005-05/msg00204.html>

- *From:* lee <lee@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx>
 - *Date:* Mon, 9 May 2005 12:20:06 -0700
-

Pretty much disabled the transaction log completely. We are using BULK INSERT to populate the data. All other approaches are just to painfully slow. Also, for all those who don't know (like I didn't), SQL server prefers smaller BULK INSERT files rather than 1 big one (everything is relative, small for me is under 40 megs and large is above 500 megs).

For the Data file not growing, it's difficult not to grow it. I've started it at 5 GB but it only takes it about 30 minutes to get passed that. I've got it now to approximately 185 Megs of data a minute. The big difference in speed comes from doing smaller files rather than 1 big file.

ORACLE seems to handle that 1 big file really well but with small files SQL server seems to only be about 1/3 slower then the oracle.

Need to figure out how to improve the time it takes to generate that INDEX

Hardware:

APP SERVER RENDERING SQL FILES

HD are SCSI 15K Seagate cheetahs
Dual XEON 3 GHZ
2 GIGS RAM

SQL SERVER

HD are SCSI 15K Seagate cheetahs
Dual XEON 3 GHZ
2 GIGS RAM

Can't run this process on 1 machine, app and SQL fight for resources to much

"Allan Mitchell" wrote:

> You do not want your datafile growing at all. That will hit you and hit you hard.

Re: Generating one table with a terabyte of data

>
> No indexes, triggers anything on the table (some actually say that minimal indexes are better in 2000 than 7.0).
>
> If it slows down to a crawl then it looks as though you are losing a resource. Memory?, Processor?
>
> Have you tried batches?
>
> Instead of firing up DTS have you used BULK INSERT (the TSQL version)?
>
> Make sure your log can take it as well.
>
> Keep your log on a seperate drive to the datafile.
>
> Fast disks and a good raid config.
>
>
>
> --
>
> Allan Mitchell MCSE,MCDBA, (Microsoft SQL Server MVP)
> www.SQLDTS.com – The site for all your DTS needs.
> www.SQLIS.com – SQL Server 2005 Integration Services.
> www.Konesans.com
>
>
> "lee" <lee@xxxxxxxxxxxxxxxxxxxxxxxxxxxx> wrote in message
news:B772A8C2-F009-4E29-816C-5B6C2F45151E@xxxxxxxxxxxxxxxxxxxx
>> Hi All,
>>
>> I need ideas on how to improve a process that can generate a terabyte of
>> data in a week. Our company produces software that analyzes code and one of
>> the things we do is Code Duplication. We're processing hundreds of millions
>> of lines against each other and generate gigabytes of results. In reality we
>> generate over 20 gigabytes of data a day. We are noticing major slow downs in
>> the system. Here's how we're going about it
>>
>> 1. Non-Indexed table residing on it's own FileGroup (which is not on the
>> same drive as the SQL server).
>> 2. A process that generates Text files that we use BULK INSERT to load the
>> data (any other way of doing inserts is just to slow).
>>
>> This process should generate 10 million rows in the table per hour. Only
>> problem is that we're noticing a considerable slow down as we pummel the
>> Database. We're making sure our Bulk Insert has exclusive table locks and
>> nobody else accesses the system while this is happening. We're also noticing
>> a strange behavior at the OS level. The system just slows down to a crawl.
>> Even mouse movement becomes jagged (Win2K Server).
>>
>> I should also mention that once we're done inserting we create indexes at
>> the end. Average time to create 1 indexed field (on a different filegroup) on

Re: Generating one table with a terabyte of data

>> 100 million rows is about 3 to 4 hours. Seems a little excessive to me, must
>> be doing something wrong.
>>
>> I've gone over this 100 times and I can't see anything else to improve
>> performance. I have a friend that is an ORACLE DBA and he keeps telling me
>> that SQL server just can't handle this type of volume and that I should move
>> to ORACLE. I refuse to believe ORACLE can do more then SQL Server. He
>> actually took one of our 500 meg text files and loaded it in 1/10 the time it
>> took SQL Server. My answer to that is "I must be configuring it incorrectly".
>>
>> I've locked SQL server to 1/3 physical memory usage. DB growth is 512Meg
>> chunks. We don't need any transaction logs, so we're using RECOVERY MODEL =
>> SIMPLE (any other mode kills SQL server anyway, LOG fills to fast with our
>> data creation speed). I've thrown everything at this except the kitchen sink.
>>
>> Any suggestions on how I can improve my Bulk Inserts and how to improve my
>> Index creations? Should I consider SQL Server 2005? Will it handle my needs
>> better?
>>
>
>
>
>
.

• *References:*

- ◆ ***Generating one table with a terabyte of data***

◇ *From:* lee

- ◆ ***Re: Generating one table with a terabyte of data***

◇ *From:* Allan Mitchell

- Prev by Date: ***Re: DTS Tasks running in job random failure***
- Next by Date: ***Re: DTS Tasks running in job random failure***
- Previous by thread: ***Re: Generating one table with a terabyte of data***
- Next by thread: ***Re: Generating one table with a terabyte of data***
- Index(es):
 - ◆ ***Date***
 - ◆ ***Thread***