The Official Google Blog
Insights from Googlers into our products, technology, and the Google culture.

Sorting 1PB with MapReduce

11/21/2008 04:55:00 PM

At Google we are fanatical about organizing the world's information. As a result, we spend a lot of time finding better ways to sort information using MapReduce, a key component of our software infrastructure that allows us to run multiple processes simultaneously. MapReduce is a perfect solution for many of the computations we run daily, due in large part to its simplicity, applicability to a wide range of real-world computing tasks, and natural translation to highly scalable distributed implementations that harness the power of thousands of computers.

In our sorting experiments we have followed the rules of a standard terabyte (TB) sort benchmark. Standardized experiments help us understand and compare the benefits of various technologies and also add a competitive spirit. You can think of it as an Olympic event for computations. By pushing the boundaries of these types of programs, we learn about the limitations of current technologies as well as the lessons useful in designing next generation computing platforms. This, in turn, should help everyone have faster access to higher-quality information.

We are excited to announce we were able to sort 1TB (stored on the Google File System as 10 billion 100-byte records in uncompressed text files) on 1,000 computers in 68 seconds. By comparison, the previous 1TB sorting record is 209 seconds on 910 computers.

Sometimes you need to sort more than a terabyte, so we were curious to find out what happens when you sort more and gave one petabyte (PB) a try. One petabyte is a thousand terabytes, or, to put this amount in perspective, it is 12 times the amount of archived web data in the U.S. Library of Congress as of May 2008. In comparison, consider that the aggregate size of data processed by all instances of MapReduce at Google was on average 20PB per day in January 2008.

It took six hours and two minutes to sort 1PB (10 trillion 100-byte records) on 4,000 computers. We're not aware of any other sorting experiment at this scale and are obviously very excited to be able to process so much data so quickly.

An interesting question came up while running experiments at such a scale: Where do you put 1PB of sorted data? We were writing it to 48,000 hard drives (we did not use the full capacity of these disks, though), and every time we ran our sort, at least one of our disks managed to break (this is not surprising at all given the duration of the test, the number of disks involved, and the expected lifetime of hard disks). To make sure we kept our sorted petabyte safe, we asked the Google File System to write three copies of each file to three different disks.

Significantly improved handling of the so-called "stragglers" (parts of computation that run slower than expected) was a key software technique that helped sort 1PB. And of course, there are many other factors that contributed to the result. We'll be discussing all of this and more in an upcoming publication. And you can also check out the video from our recent Technology RoundTable Series.

Posted by Grzegorz Czajkowski, Systems Infrastructure Team

Permalink

Links to this post

…und das Leben nach SQL geht weiter … jetzt wird reduziert!
blog carnival: google
hadoop sorts a petabyte in 16.25 hours and a terabyte in 62 seconds
history of mapreduce, part 2
dewitt and stonebraker vs. mapreduce, round 2
terasort
где мои пакеты, чувак?
map reduce sort benchmark
links for 2008-12-12
wanted: hard drive boys for our new ginormous data center
google sorts gobs of data really fast, people freak out
-all the news i missed
google breaks speed record for processing data online
google sorts 1tb in 68 secondes and 1pb in …
[science][database][google]google が 1pb のデータを mapreduce でソート
sorting data
meneados que llegaron a portada
the morning brew #231
data scales
новости google: галлопом по америкам
sorting 1pb with mapraduce at google
qcon san francisco 2008 - day 3 (alas, the last one)
[google] google puede ordenar datos del 1pb(petabyte) de datos en ...
谷歌宣布：处理1tb数据只需68秒1pb六个小时
weekly ten (11-24-2008)
la tecnologia di google
google sorting 1pb with mapreduce
stumdoma paieška ir kitos google naujienos
сортировка петабайта данных заняла у google шесть часов две минуты
mapreduce sorts 1pb of data in 6 hours.
mapreduce sorts 1pb of data in 6 hours.
google anuncia que puede ordenar un petabyte en 6 horas y 2 minutos
google 每天处理约20000tb 的数据
google sorts 1 petabyte in 6 hours
google이 1pb자료의 정렬을 mapreduce를 사용하여 성공했다고 합니다
just for fun: google
vitor magalhães: just for fun: google
googleは1pbのデータを6時間でソートできる
getting things in order quickly
links for 2008-11-23
ein petabyte sortieren: 6:20h

google: try this on windows server
how to sort 1tb of data in 68 seconds or even 1 pb
google trie 1 petabytes de données !
google mapreduce wins terasort
google sorted 1pb with mapreduce
google が 1pb のデータを mapreduce でソート
official google blog: třídění 1 petabytu s mapreduce
busy? mind map!
google: sorting 1pb with mapreduce
sorting a petabyte
large-scale translation memories (google)
google can sort
google architecture
google 每天处理约20000tb 的数据

Create a Link

Newer Post Home Older Post