

Forum:[Hack Forums](#)

Thread: [Best way to do a large database search.](#)

ThisBearScat - October 1, 2019, 3:45 am

Hi all,

I have an extremely large database that i have built up over time, converted to a txt file and removed duplicates yet the size is about ***GB. i have tried many ways to speed up searching the DB. i currently have split the file and search using grep but with the size of it, the time a search takes is ridiculous.

would it be worth converting back to a database file?
is there any faster ways, i have heard people mention sift and B tree but have no idea how to implement/use these. i just need a quicker way to perform searches and i dont mind reorganizing/optimizing the data to make it possible.

any advice or ideas are appreciated and thanks in advance.

pein - October 1, 2019, 3:59 am

That's huge, what are you using it for? Like in general? What you search for?

ThisBearScat - October 1, 2019, 4:05 am

(Oct 1, 2019 03:59 AM)pein Wrote: [\[CITE\]](#) That's huge, what are you using it for? Like in general? What you search for?

mainly i use it for investigating as it generates more leads. for instance if a username is linked to multiple emails or vice versa then it gives you more to look into

compalabee - October 3, 2019, 6:41 am

By the sound of it, since you're talking GB's worth of data, you might want to invest on a RAM and SSD instead of a programmatic way to fix this. In my experience, the standard 4GB laptop (even on a modular OS like Linux or whatever you got) would start to slow down fast. You can try to tinker with your swap file and see if other sizes would be more viable, but otherwise, a RAM is always best.

After that, just cat and grep, done deal.

ThisBearScat - October 3, 2019, 7:03 am

(Oct 3, 2019 06:41 AM)compalabee Wrote: [\[CITE\]](#) By the sound of it, since you're talking GB's worth of data, you might want to invest on a RAM and SSD instead of a programmatic way to fix this. In my experience, the standard 4GB laptop (even on a modular OS like Linux or whatever you got) would start to slow down fast. You can try to tinker with your swap file and see if other sizes would be more viable, but otherwise, a RAM is always best.

After that, just cat and grep, done deal.

thanks for the advise, the system has 16gb ram and solid state drives but grep just seems slow with the amount of data, maybe i should look to try and utilize the cloud maybe. i have also heard that its possible to use multiple servers to spread the data to speed up searches but this seems like an expensive option to something that probably has a better solution.

BAHLuypGqU2 - *October 4, 2019, 5:14 am*

I have an 18TB dataset that is deduped, I use a full-text indexing program (that has its own 2TB index), but for searches on hashes, sites, emails, userids, birthdates etc I just run `yara > grep` with a quick custom script that auto generates a yara index file then a grep command-based of inputs. With all the dumps that come out, it's too much time investment to clean them up. I recommend a NAS with gigabit connection and ssh in run x script and profit mate.

P.S don't cloud it, I tried it costs way too much!