

Forum:[Hack Forums](#)

Thread: [Import huge list into mysql database](#)

RayGenWurm - February 10, 2019, 5:49 am

Hey everyone,

i got my hands on "collection #1" password list.

since the collection is split into hundreds of files i am looking for a way to search trough all of them.

My idea at the moment is to import all lists into a mysql database and then write a simple script to search inside the database.

My problem is im not sure whats the best way to import this huge data.
We are talking here about 100 millions of lines.

First try: write a script that parses line by line of every file and write them into the database => of course takes ages

Second try: a php script using the LOAD FILE command to load whole files into the database => works okay but takes longer the larger the database gets

Third: using "LOAD FILE" via comand line with mysql => faster but also same problem as second try

so the main problem is: The larger the database gets the longer it takes to add new data.
For example first file with 100.000 lines get imported in 8 seconds... the third file with 100.000 lines already takes 40 seconds... 10'th file with 100.000 lines takes about 6 minutes...

do you have any ideas how to import the data?

Digital - February 10, 2019, 5:51 am

Use mysql cli to import, more faster and stable than GUIs, increase mysql cnf timeouts, max file size etx

RayGenWurm - February 10, 2019, 5:55 am

(Feb 10, 2019 06:51 AM)Digital Wrote: [\[CITE\]](#) Use mysql cli to import, more faster and stable than GUIs, increase mysql cnf timeouts, max file size etx

like in sayed i already tried thie cli to import..

its way faster but it gets slower with every file i try to import

im am on default settings right now.. anything i can tweak?

i will take a look at cnf timeouts, max file size is not an issue with cli right?

Digital - February 10, 2019, 5:58 am

(Feb 10, 2019 06:55 AM)LoL-Fighter Wrote: [\[CITE\]](#) like in sayed i already tried thie cli to import..

its way faster but it gets slower with every file i try to import

im am on default settings right now.. anything i can tweak?

i will take a look at cnf timeouts, max file size is not an issue with cli right?

How about Load Data Infile? That would be useful

HFDone - February 10, 2019, 6:02 am

you better use language that can use threading to do it

+Haze - February 10, 2019, 6:27 am

Well, if you're uploading millions of files and want to be able to easily search through them, MySQL is proving to be slow and it doesn't seem like the data is relational at the moment.

Therefore, I suggest using a distributed search engine. A search engine might seem a bit overkill, but solely because you have large amounts of data, and you have a requirement to be able to retrieve them quickly, I think you'll benefit from one.

A great search engine that I know that has the capabilities to store petabytes of data and can scale very well is Elastic Search. You can search through millions of files in the blink of an eye. It has a bit of a learning curve, but in my opinion, it's the next thing to learn after you've become comfortable with relational databases.

<https://www.elastic.co/products/elasticsearch>

Alternatively, you can use MongoDB, but I am unsure of it's application as I've never used it before.

itzDecon - March 4, 2019, 3:44 pm

It seems like SQL is not the way you want to go with this, especially if you're going to have hundreds of thousands of entries within. Unless you decide to go with a language that allows threading, importing and searching this data will probably take a very long time regardless of how you decide to do it. I would also suggest something like +Haze said, a search engine could be your best bet for something of this application.

ChaoticHaze - March 4, 2019, 5:53 pm

If you really wanna do it in PHP here is a link to how to use multithreading in PHP:

<http://php.net/manual/en/book.pthreads.php>