

CLAIMS

What is claimed:

1. A method comprising:

5 generating a service querying request, the service querying request including a querying condition;

acquiring a first server address list of a target service corresponding to the querying condition;

10 presetting a failure database, the failure database storing an association relationship between one or more failed server addresses and corresponding statistical numbers of failure times, the corresponding statistical numbers of failure times being based on types of failure and used to make statistics on a number of failure times of the one or more failed server addresses within a preset time period;

15 acquiring a failed server address list of a preset number of failed server addresses selected from the one or more failed addresses that have top rankings based on respective statistical numbers of failure times of the one or more failed addresses, the preset number representing a maximum number of failed servers to be isolated, the maximum number based on a demand associated with the querying condition and being less than or equal to an actual number of failed servers;

20 deleting the failed server address list from the first server address list to obtain a second server address list;

selecting a target server address from the second server address list; and

initiating an invocation of the target service to the target server address.

25 2. The method of claim 1, wherein the deleting the failed server address list from the first server address list to obtain a second server address list includes:

deleting one or more server addresses in the failed server address list from the first server address list and

30 using one or more remaining server addresses in the first server address list for the second server address list.

3. The method of claim 1, wherein the acquiring the failed server address list includes acquiring the one or more failed server addresses matching addresses in the first server address list from the failure database.

4. The method of claim 3, wherein the acquiring the failed server address list further includes:

determining that more than one failed server address is matched with the first server address list; and

5 organizing the preset number of failed server addresses that have the top rankings in the sorting into the failed server address list.

5. The method of claim 1, further comprising:

determining that there is a failure of invocation of the target service;

10 determining the target server address as a failed server address;

determining whether the failed server address exists in the failure database; and

upon determining that the failed server address exists in the failure database, increasing a statistical number of failure times corresponding to the failed server address by a preset threshold.

15

6. The method of claim 5, further comprising:

upon determining that the failed server address does not exist in the failure database:

adding the failed server address into the failure database,

setting a corresponding statistical number of failure times at a preset initial

20 value, and

updating the failed server address list.

7. The method of claim 6, wherein the failure of the invocation of the target service includes at least one of the following:

25 a response timeout; and

a busy degree of target server thread pool is beyond a preset threshold.

8. The method of claim 1, further comprising:

respectively setting a timer for each of the failed server addresses in the failure

30 database; and

when the timer expires, resetting the statistical number of failure times corresponding to the corresponding failed server address to zero.

9. The method of claim 1, wherein the acquiring the failed server address list includes:

generating a failure information acquiring request, the failure information acquiring request acquiring information of a server that has ever been failed in a cluster within the preset time period;

sending the failure information acquiring request to a remote server, the remote server storing a record of address information of the failed server; and

receiving the failed server address list returned by the remote server with respect to the failure information acquiring request.

10. The method of claim 1, wherein the selecting the target server address from the second server address list includes randomly selecting the target server address from the second server address list.

11. An apparatus comprising:

one or more processors; and

memory communicatively coupled to the one or more processors, the memory storing a plurality of modules including:

a request generating module that generates a service querying request, the service querying request including a querying condition;

a first address acquiring module that acquires a first server address list of a target service corresponding to the querying condition;

a database presetting module that presets a failure database, the failure database storing an association relationship between one or more failed server addresses and corresponding statistical numbers of failure times, the corresponding statistical numbers of failure times being based on types of failure and used to make statistics on a number of failure times of the one or more failed server addresses within a preset time period;

a failure address acquiring module that acquires a failed server address list of a preset number of failed server addresses selected from the one or more failed addresses that have top rankings based on respective statistical numbers of failure times of the one or more failed addresses, the preset number representing a maximum number of failed servers to be isolated, the maximum number based on a demand

associated with the querying condition and being less than or equal to an actual number of failed servers;

a second address acquiring module that deletes the failed server address list from the first server address list to obtain a second server address list;

5 an address selecting module that selects a target server address from the second server address list; and

a service invoking module that initiates an invocation of the target service to the target server address.

10 12. The apparatus of claim 11, wherein the second address acquiring module further: deletes one or more server addresses in the failed server address list from the first server address list; and

uses one or more remaining server addresses in the first server address list for the second server address list.

15

13. The apparatus of claim 11, wherein the failure address acquiring module includes:

an address matching sub-module that acquires the failed server address list includes acquiring the one or more failed server addresses matching addresses in the first server
20 address list from the failure database.

14. The apparatus of claim 13, wherein the failure address acquiring module further includes:

25 an address sorting sub-module that determines that more than one failed server address is matched with the first server address list; and

a list organizing sub-module that organizes the preset number of failed server addresses that have the top rankings in the sorting into the failed server address list.

15. The apparatus of claim 11, wherein the plurality of modules further includes:

30 a determining module that determines that there is a failure of invocation of the target service and determines the target server address as a failed server address;

a judging module that determines that the failed server address exists in the failure database; and

a first calculating module that increases a statistical number of failure times corresponding to the failed server address by a preset threshold.

16. The apparatus of claim 11, wherein the plurality of modules further includes:

5 a determining module that determines that there is a failure of the invocation of the target service and determines the target server address as a failed server address;

a judging module that determines that the failed server address does not exist in the failure database;

10 a second calculating module that adds the failed server address into the failure database and sets a corresponding statistical number of failure times at a preset initial value; and

a list updating module that updates the failed server address list.

15 17. The apparatus of claim 11, wherein the failure address acquiring module includes:

a failure request generating sub-module that generates a failure information acquiring request, the failure information acquiring request acquiring information of a server that has ever been failed in a cluster within the preset time period;

20 a failure request sending sub-module that sends the failure information acquiring request to a remote server, the remote server storing a record of address information of the failed server; and

a failure information receiving module that receives the failed server address list returned by the remote server with respect to the failure information acquiring request.

25 18. One or more memories having stored thereon computer-executable instructions that are executable by one or more processors to perform operations comprising:

generating a service querying request, the service querying request including a querying condition;

30 acquiring a first server address list of a target service corresponding to the querying condition;

presetting a failure database, the failure database storing an association relationship between one or more failed server addresses and corresponding statistical numbers of failure times, the corresponding statistical numbers of failure times being based on types of failure

and used to make statistics on a number of failure times of the one or more failed server addresses within a preset time period;

acquiring a failed server address list of a preset number of failed server addresses selected from the one or more failed addresses that have top rankings based on respective statistical numbers of failure times of the one or more failed addresses, the preset number representing a maximum number of failed servers to be isolated, the maximum number based on a demand associated with the querying condition and being less than or equal to an actual number of failed servers;

deleting the failed server address list from the first server address list to obtain a second server address list;

selecting a target server address from the second server address list; and
initiating an invocation of the target service to the target server address.

19. The one or more memories of claim 18, wherein the operations further comprise:
deleting one or more server addresses in the failed server address list from the first server address list; and

using one or more remaining server addresses in the first server address list for the second server address list.

20. The one or more memories of claim 18, wherein the operation of acquiring the failed server address list further comprises:

acquiring the one or more failed server addresses matching addresses in the first server address list from the failure database;

determining that more than one failed server address is matched with the first server address list; and

organizing the preset number of failed server addresses that have the top rankings in the sorting into the failed server address list.