CLAIMS

What is claimed is:

1. A resource scheduling method, applied in an elastic-search system, the elastic-search system comprises a client proxy, a writer cluster and a reader cluster; the writer cluster includes a plurality of first nodes, the plurality of first nodes are used to write data to an index of the elastic-search system, and shards of the index are distributed in the first nodes; the reader cluster includes a plurality of second nodes, the plurality of second nodes are used to read data from the index of the elastic-search system, and replicas of the shards of the index are distributed in the second nodes; wherein the method comprises:

receiving, by the client proxy, a data write request and a data read request, wherein the data write request is used to write data into the index, and the data read request is used to read data in the index;

distributing, by the client proxy, the data write request to the writer cluster;

processing, by a first node of the writer cluster, the data write request by writing data into one or more shards distributed in the first node;

distributing, by the client proxy, the data read request to the reader cluster; and

processing, by a second node of the reader cluster, the data read request by reading data from one or more replicas distributed in the second node.

2. The method according to claim 1, wherein:

a quantity of resources occupied by one or more shards of the index are adjustable in the writer cluster; and/or,

a quantity of resources occupied by one or more replicas of the index are adjustable in the reader cluster.

3. The method according to claim 2,

wherein the quantity of resources occupied by the one or more shards of the index are adjustable in the writer cluster, by: adjusting a node specification of one or more first nodes in the writer cluster, wherein the node specification of the first node comprises a quantity of processor cores included in the first node and/or a memory size of the first node; and/or,

wherein the quantity of resources occupied by the one or more replicas of the index are adjustable in the reader cluster, by: adjusting a node specification of one or more second nodes in the reader cluster, wherein the node specification of the second node comprises a quantity of processor cores included in the second node and/or a memory size of the second node.

4. The method according to claim 2 or 3, wherein the quantity of resources occupied by the one or more shards of the index are adjustable in the writer cluster, by:

adding one or more new first nodes to the writer cluster; or,

deleting one or more first nodes in the writer cluster.

5. The method according to claim 4, wherein the method comprises:

after the one or more new first nodes are added to the writer cluster, migrating one or more shards of the index from one or more old first nodes to the one or more new first nodes in the writer cluster; or,

before the one or more first nodes are deleted in the writer cluster, migrating one or more shards of the index from one or more to-be-deleted first nodes to other first nodes in the writer cluster.

6. The method according to claim 5, wherein a policy of migrating a shard in the writer cluster is a load balancing policy.

7. The method according to claim 5 or 6, wherein the step of migrating the one or more shards of the index from the one or more old first nodes to the one or more new first nodes in the writer cluster, comprises:

migrating a target shard of the index from the old first node to the new first node, wherein the target shard satisfes at least one of the following conditions:

an amount of data written into the target shard is less than a first threshold; and

a total quantity of shards distributed in the old first node in which the target shard is located is greater than a second threshold.

8. The method according to any one of claims 2 to 7, wherein the quantity of resources occupied by the one or more replicas of the index are adjustable in the reader cluster, by:

adding one or more new second nodes to the reader cluster, and/or, adding one or more new replicas of the index in the reader cluster; or,

deleting one or more second nodes in the reader cluster, and/or, deleting one or more replicas of the index in the reader cluster.

9. The method according to claim 8, wherein the method comprises:

after the one or more new second nodes are added to the reader cluster, migrating one or more replicas of the index from one or more old second nodes to the one or more new second nodes in the reader cluster, and/or, adding one or more new replicas of the index on the added second node; or,

before the one or more second nodes are deleted in the reader cluster, migrating one or more replicas of the index from one or more to-be-deleted second nodes to other second nodes in the reader cluster, or, deleting one or more replicas in the one or more to-be-deleted second nodes.

10. The method according to any one of claims 2 to 9,

wherein the quantity of resources occupied by the one or more shards of the index are

adjustable in the writer cluster, by: adjusting, when a data write rate of the index falls outside a target rate range, the quantity of resources occupied by the one or more shards of the index; and/or,

the quantity of resources occupied by the one or more replicas of the index are adjustable in the reader cluster, by: adjusting, when a data read frequency of the index falls outside the target frequency range, the quantity of resources occupied by the one or more replicas of the index.

11. an elastic-search system, wherein the elastic-search system comprises a client proxy, a writer cluster and a reader cluster;

the writer cluster includes a plurality of first nodes, the plurality of first nodes are used to write data to an index of the elastic-search system, and shards of the index are distributed in the first nodes;

the reader cluster includes a plurality of second nodes, the plurality of second nodes are used to read data from the index of the elastic-search system, replicas of the shards of the index are distributed in the second nodes;

the client proxy is configured to receive a data write request and a data read request, distribute the data write request to the writer cluster, and distribute the data read request to the reader cluster, wherein the data write request is used to write data into the index, and the data read request is used to read data in the index;

a first node of the writer cluster is configured to process the data write request by writing data into one or more shards distributed in the first node; and

a second node of the reader cluster is configured to process the data read request by reading data from one or more replicas distributed in the second node.

12. The elastic-search system according to claim 11, wherein:

a quantity of resources occupied by one or more shards of the index are adjustable in the writer cluster; and/or,

a quantity of resources occupied by one or more replicas of the index are adjustable in the reader cluster.

13. The elastic-search system according to claim 12, wherein the elastic-search system comprises a resource scheduling apparatus, and the resource scheduling apparatus is configured to:

adjust a node specification of one or more first nodes in the writer cluster, wherein the node specification of the first node comprises a quantity of processor cores included in the first node and/or a memory size of the first node; and/or,

adjust a node specification of one or more second nodes in the reader cluster, wherein the node specification of the second node comprises a quantity of processor cores included in the

second node and/or a memory size of the second node.

14. The elastic-search system according to claim 12 or 13, wherein the elastic-search system comprises a resource scheduling apparatus, and the resource scheduling apparatus is configured to:

add one or more new first nodes to the writer cluster; or,

delete one or more first nodes in the writer cluster.

15. The elastic-search system according to claim 14, wherein the resource scheduling apparatus is configured to:

after the one or more new first nodes are added to the writer cluster, migrate one or more shards of the index from one or more old first nodes to the one or more new first nodes in the writer cluster; or,

before the one or more first nodes are deleted in the writer cluster, migrate one or more shards of the index from one or more to-be-deleted first nodes to other first nodes in the writer cluster.

16. The elastic-search system according to claim 15, wherein a policy of migrating a shard in the writer cluster is a load balancing policy.

17. The elastic-search system according to claim 15 or 16, wherein the resource scheduling apparatus is configured to migrate one or more shards of the index from one or more old first nodes to the one or more new first nodes in the writer cluster, by:

migrate a target shard of the index from the old first node to the new first node, wherein the target shard satisfes at least one of the following conditions:

an amount of data written into the target shard is less than a first threshold; and

a total quantity of shards distributed in the old first node in which the target shard is located is greater than a second threshold.

18. The elastic-search system according to any one of claims 12 to 17, wherein the elastic-search system comprises a resource scheduling apparatus, and the resource scheduling apparatus is configured to:

add one or more new second nodes to the reader cluster, and/or, add one or more new replicas of the index in the reader cluster; or,

delete one or more second nodes in the reader cluster, and/or, delete one or more replicas of the index in the reader cluster.

19. The elastic-search system according to claim 18, wherein the elastic-search system comprises a resource scheduling apparatus, and the resource scheduling apparatus is configured to:

after the one or more new second nodes are added to the reader cluster, migrate one or more replicas of the index from one or more old second nodes to the one or more new second nodes in

the reader cluster, and/or, add one or more new replicas of the index on the added second node; or,

before the one or more second nodes are deleted in the reader cluster, migrate one or more replicas of the index from one or more to-be-deleted second nodes to other second nodes in the reader cluster, or, delete one or more replicas in the one or more to-be-deleted second nodes.

20. The elastic-search system according to any one of claims 12 to 19, wherein the elastic-search system comprises a resource scheduling apparatus, and the resource scheduling apparatus is configured to:

adjust, when a data write rate of the index falls outside a target rate range, the quantity of resources occupied by the one or more shards of the index; and/or,

adjust, when a data read frequency of the index falls outside the target frequency range, the quantity of resources occupied by the one or more replicas of the index.

21. A computing device cluster, comprising at least one computing device, wherein each computing device comprises a processor and a memory; and

the processor in the at least one computing device is configured to execute instructions stored in the memory in the at least one computing device, so that the computing device cluster performs the method according to any one of claims 1 to 10.

22. A computer-readable storage medium, wherein the computer-readable storage medium stores computer program instructions, and when the computer program instructions are executed by a computing device cluster, the computing device cluster performs the method according to any one of claims 1 to 10.

23. A computer program product comprising instructions, wherein when the instructions are run by a computing device cluster, the computing device cluster is enabled to perform the method according to any one of claims 1 to 10.