

CLAIMS:

1. A supercluster, comprising:
a plurality of machines comprising a first machine, a second machine, and a third machine; and
a publish-subscribe (Pub-Sub) channel to which each of the plurality of machines is configurable to be subscribed,
wherein each of the plurality of machines is configured to:
publish availability status and resource allocation information to the Pub-Sub channel;
receive availability status and resource allocation information of at least another one of the plurality of machines via the Pub-Sub channel;
forward a request received from a client for content data to the another one of the plurality of machines based on the availability status and the resource allocation information on the Pub-Sub channel, and
wherein the first machine is configured to own the content data owned by the second machine based on the availability status and the resource allocation information on the Pub-Sub channel and based on the second machine authorizing the first machine to own the content data, and wherein the first machine is configured to provide the content data to the third machine based on receiving a relayed request for the content data.
2. The supercluster of claim 1, wherein
the availability status indicates whether each of the plurality of machines is available to store content data; and
the resource allocation information identifies the content data stored by each of the plurality of machines.
3. The supercluster of claim 1 or 2, wherein the Pub-Sub channel comprises a multicast address.

4. The supercluster of any one of claims 1 to 3, wherein forwarding the request received from the client for the content data to the another one of the plurality of machines based on the availability status and the resource allocation information on the Pub-Sub channel comprises:

receive, by a first machine of the plurality of machines, the request for the content data from the client;

determine, by the first machine, that the first machine does not store the content data;

determine, based on the availability status and the resource allocation information on the Pub-Sub channel, that the content data is stored on a second machine of the plurality of machines; and

forward the request to the second machine.

5. The supercluster of any one of claims 1 to 4, wherein, to own the content data currently owned by the second machine, the first machine is configured to:

identify, based on the availability status and the resource allocation information on the Pub-Sub channel, the content data owned by the second machine;

publish, to the Pub-Sub channel, availability to own the content data owned by the second machine; and

own the content data owned by the second machine.

6. The supercluster of any one of claims 1 to 5, wherein, to own the content data currently owned by the second machine, the first machine is configured to:

determine, based on the availability status and the resource allocation information on the Pub-Sub channel, that a number of second machines currently storing the content data is below a threshold;

publish, to the Pub-Sub channel, availability to own the content data;

receive, via the Pub-Sub channel, authorization from the second machines currently storing the content data;

own the content data;

receive, from a third machine of the plurality of machines, a relayed request for the content data; and

provide the content data in response to the relayed request.

7. The supercluster of any one of claims 1 to 6, wherein a first machine of the plurality of machines is configured to establish a bridge to a second machine in another supercluster, wherein:

the another supercluster neighbors the supercluster; and

the bridge is configured to relay, to the first machine, an availability status and resource allocation information of each of at least some machines of the another supercluster.

8. The supercluster of claim 7, wherein:

the first machine is configured to determine, based on the availability status of the plurality of machines on the Pub-Sub channel, that a number of unavailable machines of the plurality of machines exceeds a threshold; and

the supercluster is configured to establish the bridge in response to determining that the number of unavailable machines exceeds the threshold.

9. The supercluster of claim 7 or 8, wherein the bridge comprises a Transmission Control Protocol (TCP) connection.

10. The supercluster of any one of claims 7 to 9, wherein the first machine is configured to identify the other supercluster based on geographical proximity to the supercluster.

11. The supercluster of any one of claims 7 to 10, wherein the first machine is configured to identify that the other supercluster is adjacent to the supercluster by making a Domain Name System (DNS) request.

12. The supercluster of any one of claims 7 to 11, wherein the first machine is configured to publish the availability status and the resource allocation information of the at least some machines of the other supercluster on the Pub-Sub channel.

13. The supercluster of claim 12, wherein a third machine of the supercluster is configured to:

receive the request from the client for the content data;

determine, based on the resource allocation information of the at least some machines of the another supercluster on the Pub-Sub channel, that the content data is stored by a fourth machine of the another supercluster; and

forward the request to the fourth machine.

14. A method for distributing loads across a plurality of machines of a supercluster, wherein the plurality of machines comprises a first machine, a second machine, and a third machine, the method comprising:

publishing, by each of the plurality of machines of the supercluster, availability status and resource allocation information to a publish-subscribe (Pub-Sub) channel, wherein each of the plurality of machines is subscribed to the Pub-Sub channel;

receiving availability status and resource allocation information of at least another one of the plurality of machines via the Pub-Sub channel; and

forwarding a request received from a client for content data to the another one of the plurality of machines based on the availability status and the resource allocation information on the Pub-Sub channel, wherein ownership by the first machine of the content data owned by the second machine is based on the availability status and the resource allocation information on the Pub-Sub channel and based on the second machine authorizing the first machine to own the content data, and wherein the first machine provides the content data to the third machine based on receiving a relayed request for the content data.

15. A non-transitory computer-readable medium having computer-readable instructions such that, when executed, cause a processor of each of a plurality of machines of a

supercluster, wherein the plurality of machines comprises a first machine, a second machine, and a third machine, to:

publish availability status and resource allocation information to a publish-subscribe (Pub-Sub) channel, wherein each of the plurality of machines is subscribed to the Pub-Sub channel;

receive availability status and resource allocation information of at least another one of the plurality of machines via the Pub-Sub channel; and

forward a request received from a client for content data to the another one of the plurality of machines based on the availability status and the resource allocation information on the Pub-Sub channel, wherein ownership by the first machine of the content data owned by the second machine is based on the availability status and the resource allocation information on the Pub-Sub channel and based on the second machine authorizing the first machine to own the content data, and wherein the first machine provides the content data to the third machine based on receiving a relayed request for the content data.

16. A method for distributing loads across superclusters, comprising:

determining, by a first machine of a first supercluster, that a number of unavailable machines in the first supercluster exceeds a threshold;

establishing, by the first machine, a bridge with a second machine in a second supercluster, wherein the second supercluster is adjacent to the first supercluster;

receiving, by the first machine via the bridge, availability status and resource allocation information of each of a plurality of machines of the second supercluster; and

publishing, by the first machine on a publish-subscribe (Pub-Sub) channel of the first supercluster, the availability status and the resource allocation information of each of the plurality of machines of the second supercluster, wherein ownership by the first machine of the content data owned by a second machine of the second supercluster is based on the availability status and the resource allocation information on the Pub-Sub channel and based on the second machine authorizing the first machine to own the content data, and wherein the first machine provides the content data to a third machine of the first supercluster based on receiving a relayed request for the content data.

17. The method of claim 16, wherein:
the availability status indicates whether each of the plurality of machines of the second supercluster is available to store content data; and
the resource allocation information identifies the content data for which each of the plurality of machines of the second supercluster is responsible.
18. The method of claim 16 or 17, further comprising publishing, by a plurality of machines of the first supercluster on the Pub-Sub channel of the first supercluster, availability status and resource allocation information of each of the machines of the first supercluster.
19. The method of any one of claims 16 to 18, wherein the number of unavailable machines in the first supercluster is determined based on the availability status provided on the Pub-Sub channel of the first supercluster.
20. The method of any one of claims 16 to 19, wherein the Pub-Sub channel of the first supercluster comprises a multicast address.
21. The method of any one of claims 16 to 20, further comprising:
receiving, by the third machine from a client, a request for content data;
determining, by the third machine based on the availability status and the resource allocation information of each of the plurality of machines of the second supercluster on the Pub-Sub channel, that the content data is stored by a fourth machine of the second supercluster; and
forwarding the request to the fourth machine in the second supercluster.
22. The method of any one of claims 16 to 21, wherein the bridge comprises a Transmission Control Protocol (TCP) connection.
23. The method of any one of claims 16 to 22, further comprising identifying, by the first machine, the second supercluster based on geographical proximity to the first supercluster.

24. The method of any one of claims 16 to 23, wherein the first machine identifies that the second supercluster is adjacent to the first supercluster by making a Domain Name System (DNS) request.

25. A non-transitory computer-readable medium having computer-readable instructions such that, when executed, cause a processor of a first machine of a first supercluster to:

determine that a number of unavailable machines in the first supercluster exceeds a threshold;

establish a bridge with a second machine in a second supercluster, wherein the second supercluster is adjacent to the first supercluster;

receive, via the bridge, availability status and resource allocation information of each of a plurality of machines of the second supercluster; and

publish, on a publish-subscribe (Pub-Sub) channel of the first supercluster, the availability status and the resource allocation information of each of the plurality of machines of the second supercluster, wherein ownership by the first machine of the content data owned by the second machine is based on the availability status and the resource allocation information on the Pub-Sub channel and based on the second machine authorizing the first machine to own the content data, and wherein the first machine provides the content data to the third machine based on receiving a relayed request for the content data.

26. A system, comprising:

a first supercluster, wherein the first supercluster comprises:

a plurality of first machines; and

a publish-subscribe (Pub-Sub) channel to which each of the plurality of first machines is configurable to be subscribed;

a second supercluster, wherein the second supercluster comprises:

a plurality of second machines; and

a bridge between the first supercluster and the second supercluster, wherein one of the plurality of first machines is configured to:

receive, from one of the plurality of second machines via the bridge, an availability status and resource allocation information of each of the plurality of second machines; and

publish, on the Pub-Sub channel of the first supercluster, the availability status and the resource allocation information of each of the plurality of second machines, wherein the one of the plurality of first machines is configured to own the content data owned by the one of the plurality of second machines based on the availability status and the resource allocation information on the Pub-Sub channel and based on the one of the plurality of second machines authorizing the one of the plurality of first machines to own the content data, and wherein the one of the plurality of first machines provides the content data to a third machine of the first supercluster based on receiving a relayed request for the content data.