

CLAIMS

1. A method, comprising:
 - performing, by one or more computing devices:
 - determining that state transition processing operations of a state replication group (SRG) comprising a plurality of nodes are to be suspended, wherein the SRG is designated to replicate state information comprising a respective commit record set of an application, wherein each commit record of the commit record set has an associated commit sequence number (CSN) indicative of an order in which the corresponding state transition of the application was committed at the SRG;
 - identifying a target CSN up to which commit record sets of one or more nodes of the SRG are to be synchronized;
 - transmitting a respective suspend command from a configuration manager of the SRG to the one or more other nodes of the SRG, wherein the suspend command indicates the target CSN and instructs the one or more other nodes to stop processing state transition messages;
 - verifying, by a particular node of the one or more other nodes, that a commit record corresponding to the target CSN is stored in the particular node's commit record set; and
 - suspending state transition processing operations by the particular node.
2. The method as recited in claim 1, wherein said determining that the state transition processing operations of the SRG are to be suspended is responsive to a detection that a metric is outside an acceptable range, wherein the metric comprises one or more of: (a) a number of active nodes of the SRG, (b) a rate of SRG configuration-delta messages received from the configuration manager at a selected node of the SRG, or (c) a number of client connections to a selected node of the SRG.
3. The method as recited in any one of claims 1-2, wherein said determining that the state transition processing operations of the SRG are to be suspended is performed at a committer node of the SRG, wherein the committer node is responsible for committing one or

more requested state transitions of the application, and wherein the target CSN is the highest CSN among the CSNs of the commit record set of the committer node.

4. The method as recited in ~~any of~~ claims 3, further comprising sending, by the committer node to the configuration manager, a suspend request indicating the target CSN, wherein the target CSN is stored at the committer node.

5. The method as recited in any one of claims 1-4, further comprising performing, by the one or more computing devices prior to said verifying:

determining, by the particular node, in response to receiving a suspend command from the configuration manager, that the commit record set at ~~the~~ a second node does not include a commit record corresponding to the target CSN; and

requesting, by the particular node from a different node of the SRG, one or more commit

records including a commit record corresponding to the target CSN.

6. The method as recited in any one of claims 1-5, wherein ~~the particular node~~ the plurality of nodes comprises ~~a first~~ different threads of execution at a particular host, including a first thread of execution associated with the particular node, further comprising performing, by the one or more computing devices:
restarting, subsequent to said suspending, the first thread of execution.

7. The method as recited in any one of claims 1-6, further comprising performing, by the one or more computing devices:

receiving, at the configuration manager from the particular node subsequent to said verifying, a confirmation that the particular node has updated its commit record set up to the target CSN; and

including, by the configuration manager in a collection of up-to-date nodes of the SRG, the particular node.

8. The method as recited in claim 7, further comprising performing, by the one or more computing devices:

receiving, at the configuration manager subsequent to said suspending, respective messages from a second plurality of nodes of the SRG indicating that one or more of the ~~respective second plurality of~~ nodes are available for service, wherein the second plurality of nodes includes a committer node, the particular node, and a third node;

determining, by the configuration manager using the collection of up-to-date nodes, that the third node's commit record set does not include a commit record corresponding to the target CSN; and

transmitting, by the configuration manager to the third node, an indication of the target CSN.

9. The method as recited in claim 8, further comprising performing, by the one or more computing devices:

receiving, at the configuration manager, a confirmation that the third node's commit record set has been updated up to the target CSN;

determining, by the configuration manager that a number of available SRG nodes whose commit records have been updated up to the target CSN exceeds a threshold; and

transmitting, by the configuration manager to each node whose commit record sets have been updated up to the target CSN, a respective re-activation request including a representation of a targeted configuration of the SRG.

10. The method as recited in any one of claims 1-9, wherein the plurality of nodes of the SRG comprise a directed acyclic graph.

11. The method as recited in any one of claims 1-9, wherein the plurality of nodes of the SRG comprise a directed acyclic graph that includes a replication pathway from an acceptor node to a committer node, further comprising performing, by the one or more computing devices:

receiving, at the acceptor node prior to said determining that the state transition processing operations are to be suspended, a request from a client to commit a particular state transition of the application;

storing, at the acceptor node, a record indicating that the particular state transition has been accepted for replication;

propagating, from the second node via the replication pathway to the committer node, the request to commit the particular state transition;

determining, by the committer node, that a number of nodes of the SRG at which a respective record indicative of the particular state transition has been stored is above a replication threshold, and

storing, at the committer node, a commit record corresponding to the particular state transition.

12. The method as recited in any one of claims 1-11, wherein the application comprises one of: a database service, a logging service, or a control-plane component of a provider network service.

13. The method as recited in any one of claims 1-12, wherein the determining that state transition processing operations of the SRG are to be suspended is performed using a consensus protocol.

14. The method as recited in any one of claims 1-13, wherein the configuration manager of the SRG is a fault-tolerant configuration manager.

15. A system configured to implement the method of any one of claims 1-14.

16. A non-transitory, computer-readable storage medium, storing program instructions that when executed by the one or more computing devices cause the one or more computing devices to implement the method of any one of claims 1-14.

~~17. A system, comprising:~~

~~one or more computing devices configured to:~~

~~identify, at a first non-committer node of a directed acyclic graph (DAG) of nodes designated to replicate a state of an application, at least a portion of a first~~

replication pathway to a first committer node of the DAG, based at least in part on a first sequence of configuration delta messages received from a fault-tolerant DAG configuration manager, wherein at least one other node of the DAG has received a different sequence of configuration delta messages from the fault-tolerant DAG configuration manager;

transmit, from the first non-committer node, for replication at respective local storage devices of one or more nodes along the first replication pathway, a state transition message indicating (a) approval of a first candidate state transition requested by a client and (b) a sequence number indicative of an order in which the first candidate state transition is processed at the first non-committer node relative to other candidate state transitions;

in response to a determination, by the fault-tolerant DAG configuration manager, that a particular node of the first replication pathway is to be removed from the DAG,

initiate a transmission, from the fault-tolerant DAG configuration manager to one or more remaining nodes of the DAG, of a respective configuration delta message indicating a removal of the particular node from the DAG, without directing the one or more remaining nodes to pause processing of state transitions.

18.—A method, comprising:

performing, by one or more computing devices:

propagating a state transition message corresponding to a requested state transition from a first non-committer node of a directed acyclic graph (DAG) of replication nodes to a first committer node of the DAG, wherein the state transition message includes a sequence number indicative of an order in which the requested state transition was approved relative to other requested state transitions;

in response to determining, by a DAG configuration manager, that a particular node is to be added to the DAG, initiating a propagation, from the DAG configuration manager to one or more other nodes of the DAG, of a

~~configuration-delta message indicating the addition of the particular node to the DAG, without directing the one or more other nodes to pause processing of state transition messages.~~

19.—~~A non-transitory computer-accessible storage medium storing program instructions that when executed on one or more processors implement a particular node of a~~

~~DAG established to replicate state information, wherein the particular node is configured to:~~

~~receive, from a DAG configuration manager, a plurality of configuration-delta messages, wherein a particular configuration-delta message of the plurality of configuration-delta messages indicates one or more of: (a) an addition of a node to the DAG, or (b) a removal of a node from the DAG;~~

~~subsequent to receiving the particular configuration-delta message from the DAG configuration manager, indicating that an immediate predecessor node from which a state transition message was received at the particular node has been removed from the DAG,~~

~~identify, using at least the particular configuration-delta message, a different immediate predecessor node, and~~

~~transmit, to the different immediate predecessor node, a request for state transition messages.~~

20.—~~A non-transitory computer-accessible storage medium storing program instructions that when executed on one or more processors implement a configuration manager of a DAG of nodes established to replicate state information, wherein the configuration manager is configured to:~~

~~determine, based at least in part on a consensus protocol, that a particular node of the DAG is to be removed from the DAG, wherein the particular node occupies a particular position within the DAG;~~

~~remove, from a persistent representation of a configuration of the DAG stored at the configuration manager, the particular node; and~~

~~transmit, to one or more remaining nodes of the DAG, a configuration-delta message indicating that the particular node has been removed from the DAG, without~~

requesting the one or more remaining nodes to pause processing related to state transitions of the application.