MAY 16-18, 2017
MIAMI, FL

Tomcat Cluster

Keiichi Fujino

Agenda

- About me
- Tomcat Clustering Overview
- Session Replication
- Cluster Channel Component
- Monitoring Cluster components

About me

- Keiichi Fujino
- I live in Japan
- Software Enginner since 2002
- Apache Tomcat committer since 2010
  - kfujino@apache.org

Clustering Overview

What is Cluster?

- Performance improvement
- High availability

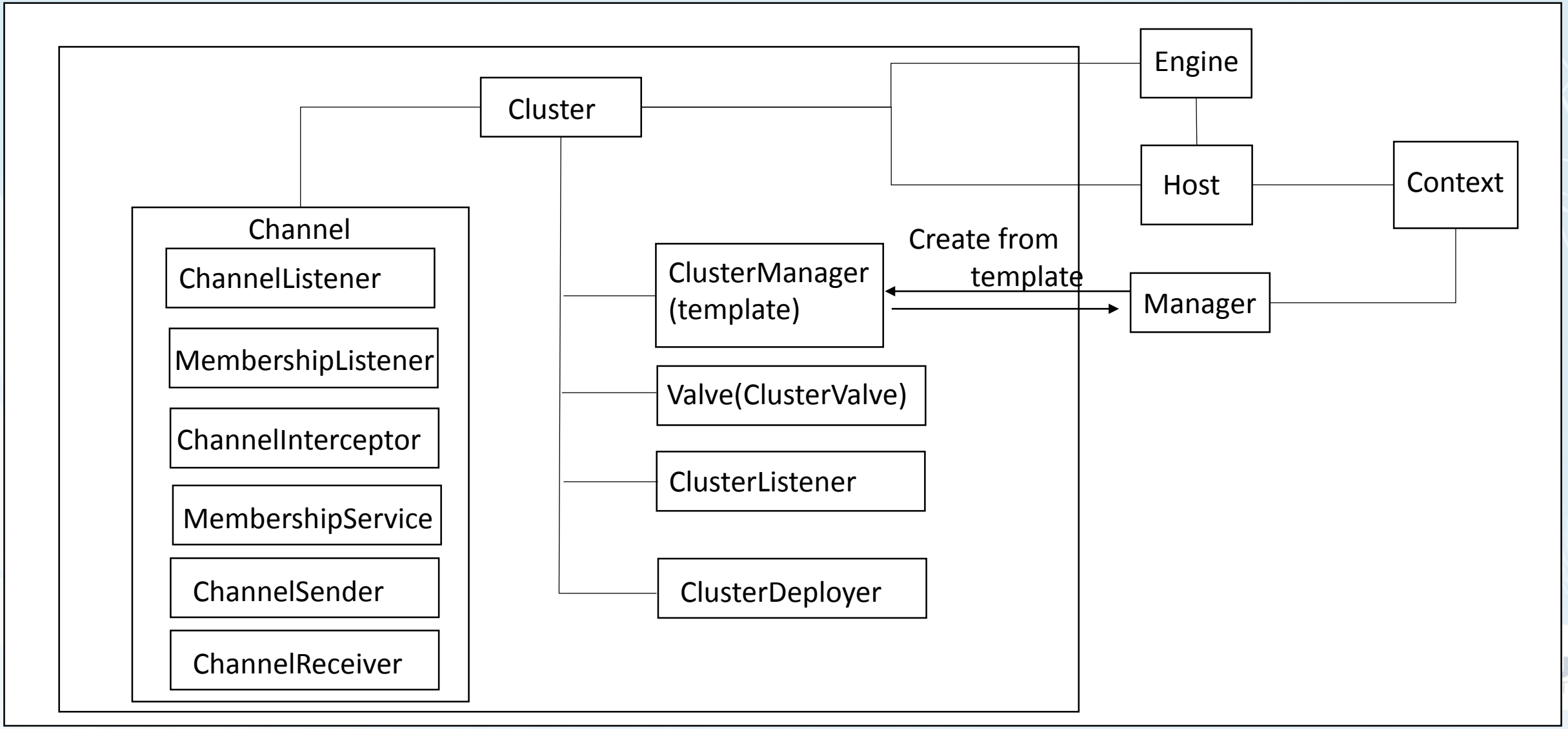
Tomcat Clustering

- Cluster membership/Grouping
- Session Replication

Load balancing is not a Tomcat features

- Use mod_jk / mod_proxy_balancer

Cluster Architecture



Cluster Architecture

- Cluster
  - The main component of Tomcat Cluster
- Cluster Manager
  - The session manager for the session replication
- Valve(Cluster Valve)
  - The same as usual Valve
  - Added to the request processing pipeline automatically

Cluster Architecture

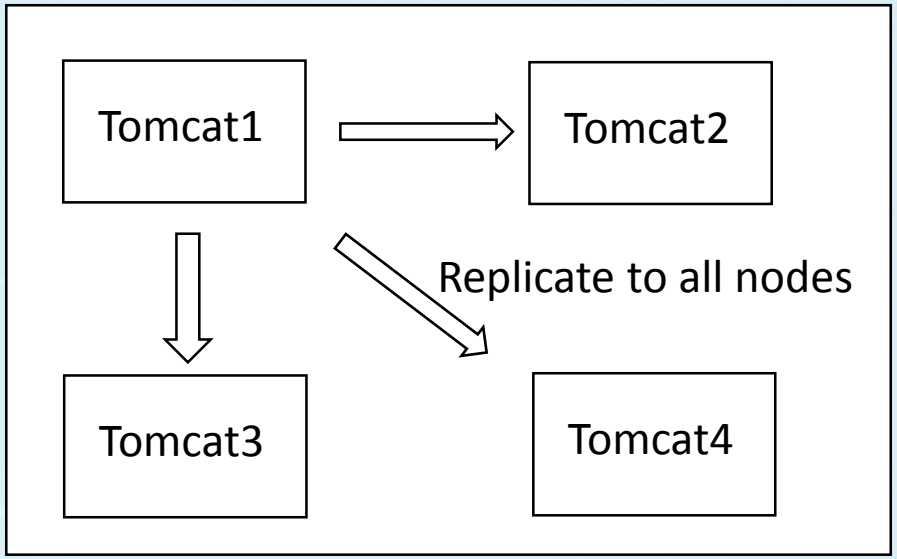
- Cluster Deployer
  - Sharing of WAR files among cluster nodes
- ClusterListener
  - Listen cluster messages and events
- Channel
  - Performs messaging and grouping among the cluster nodes

Session Replication

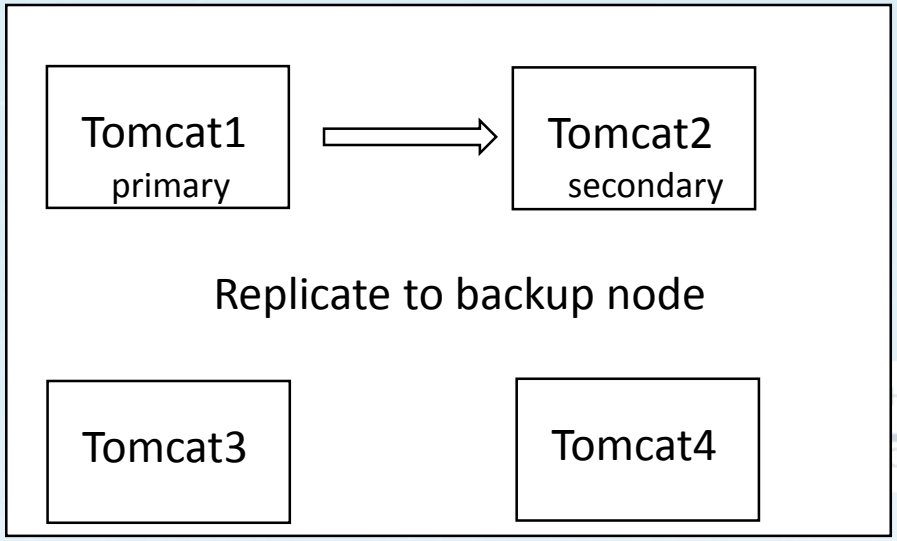
Session Replication

Implementations

- All-to-All session replication
  - DeltaManager (Default)



- Primary-Secondary session replication
  - BackupManager



Use constraints

- Make sure that your web.xml describe the <distributable/> element
- Session attributes must implement java.io.Serializable
- sticky session
  - If you use the BackupManager, This is required

How to configure

- Define <Cluster> element inside The Engine or Host element.
- Configure Cluster Manager
  - DeltaManager or BackupManager
- Configure Channel components
- Enable org.apache.catalina.ha.tcp.ReplicationValve
- Enable org.apache.catalina.ha.session.ClusterSessionListener
  - If DeltaManager use

Delta Replication

- Replicate only the changes of session
  - Not all session data

- Replicate all changes of session at the time of end of request
  - Not replication per change of session

Delta Replication

Register Delta Info

- *Add ATTRIBUTE(*Attr_A, Value_A)
- *Add ATTRIBUTE(*Attr_B, Value_B)
- *Remove ATTRIBUTE(*Attr_A)
- *Add ATTRIBUTE(*Attr_B, Value_BB)

Default : recordAllActions=false

TYPE	ACTION	NAME	VALUE
~~*ATTRIBUTE*~~	~~*SET*~~	~~Attr_A~~	~~Value_A~~
~~*ATTRIBUTE*~~	~~*SET*~~	~~Attr_B~~	~~Value_B~~
ATTRIBUTE	*REMOVE*	Attr_A	null
ATTRIBUTE	*SET*	Attr_B	Value_BB

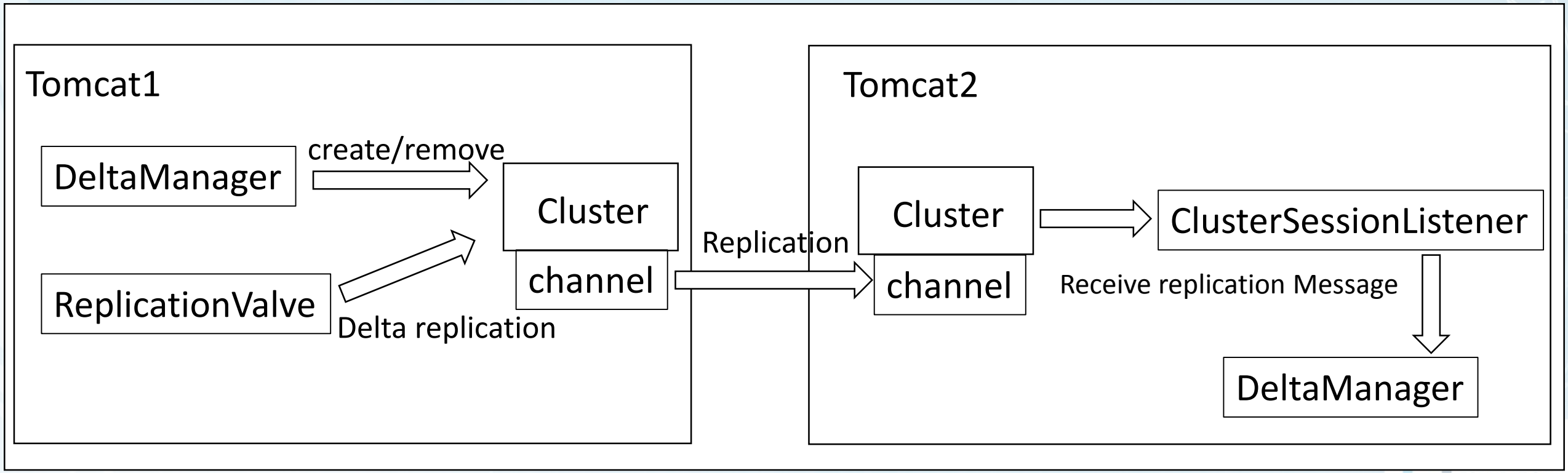
recordAllActions=true

TYPE	ACTION	NAME	VALUE
ATTRIBUTE	*SET*	Attr_A	Value_A
ATTRIBUTE	*SET*	Attr_B	Value_B
ATTRIBUTE	*REMOVE*	Attr_A	null
ATTRIBUTE	*SET*	Attr_B	Value_BB

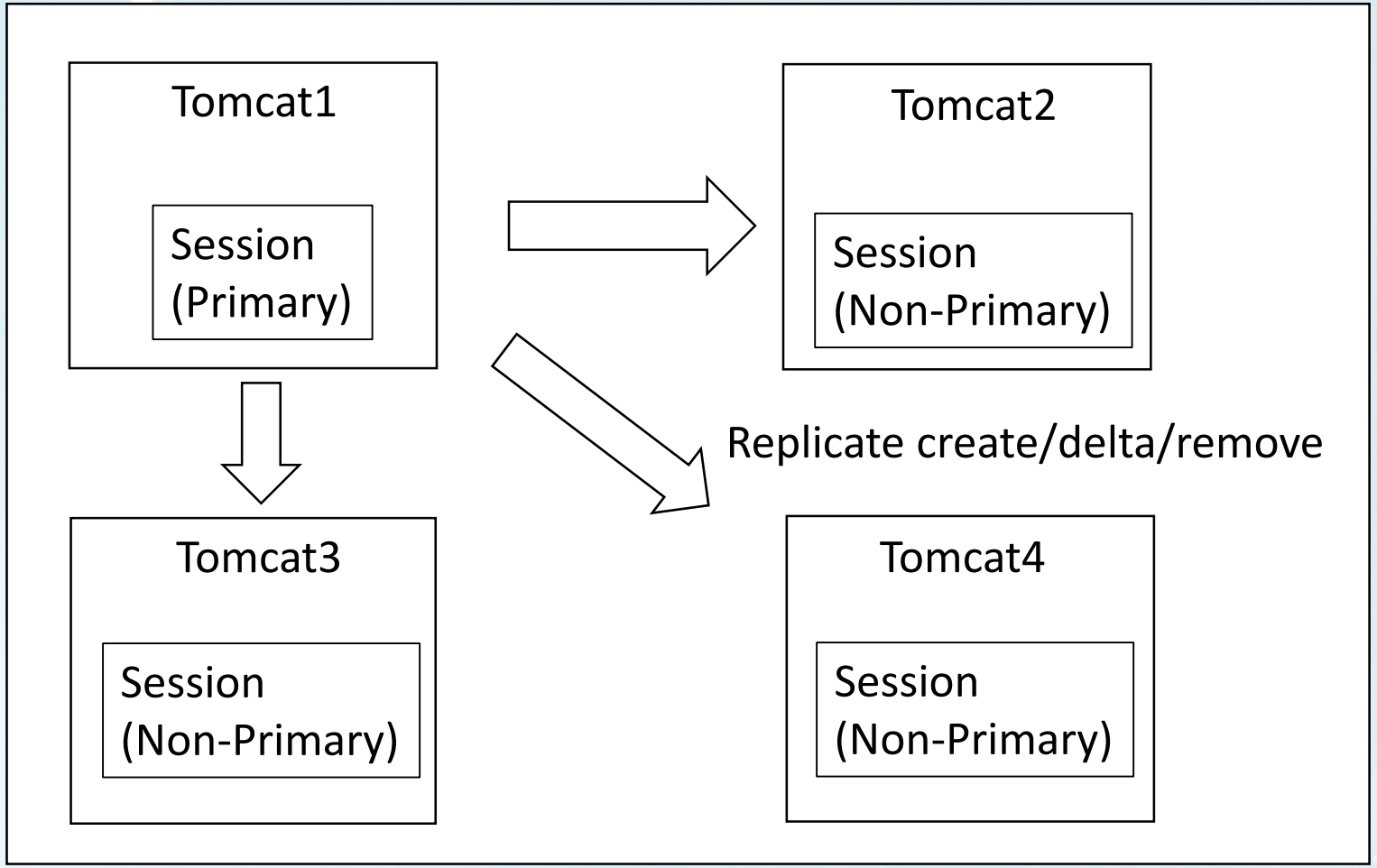
DeltaManager

- All-to-All session replication
- Default Session Manager in Cluster environment
- For small cluster

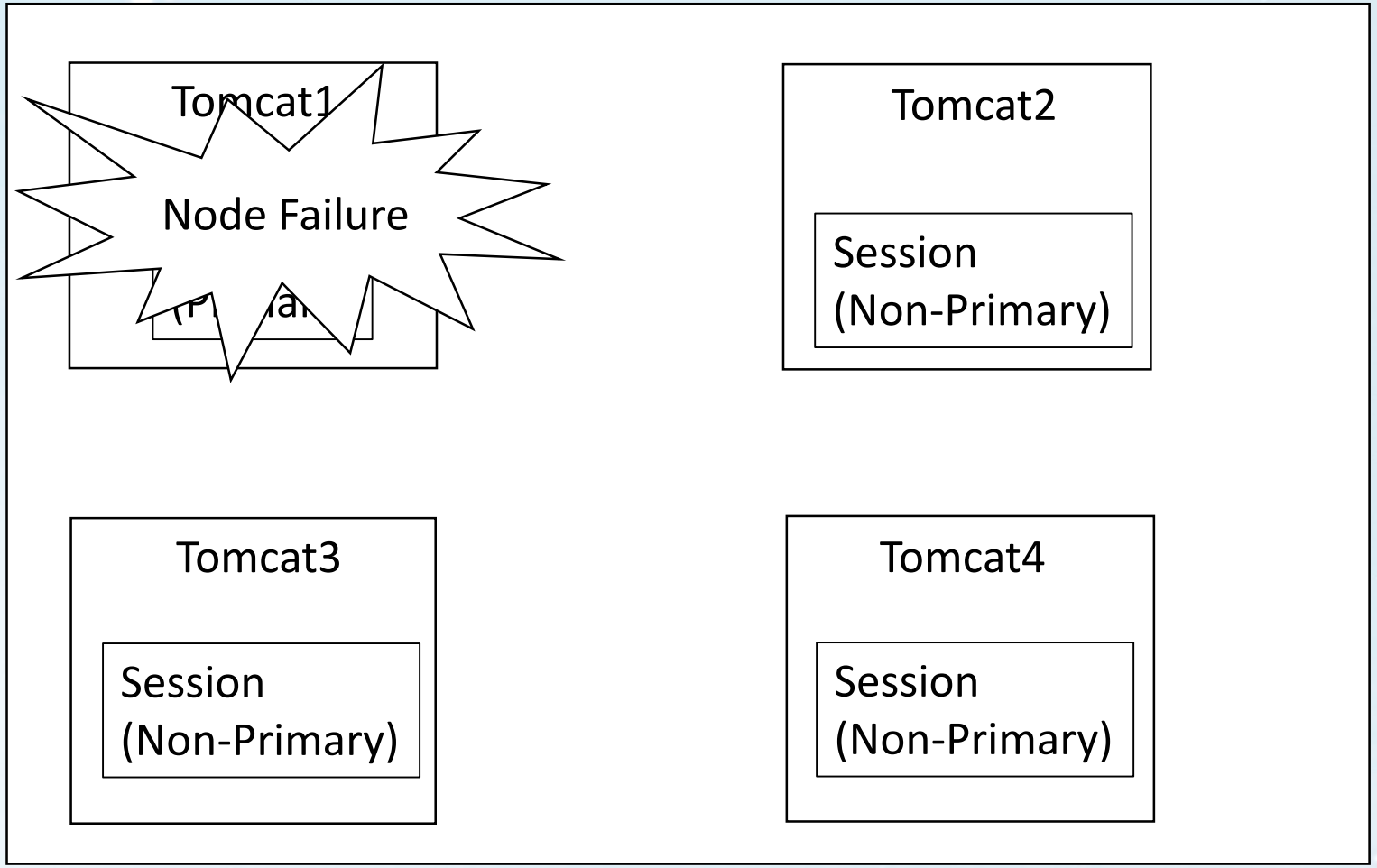
Architecture



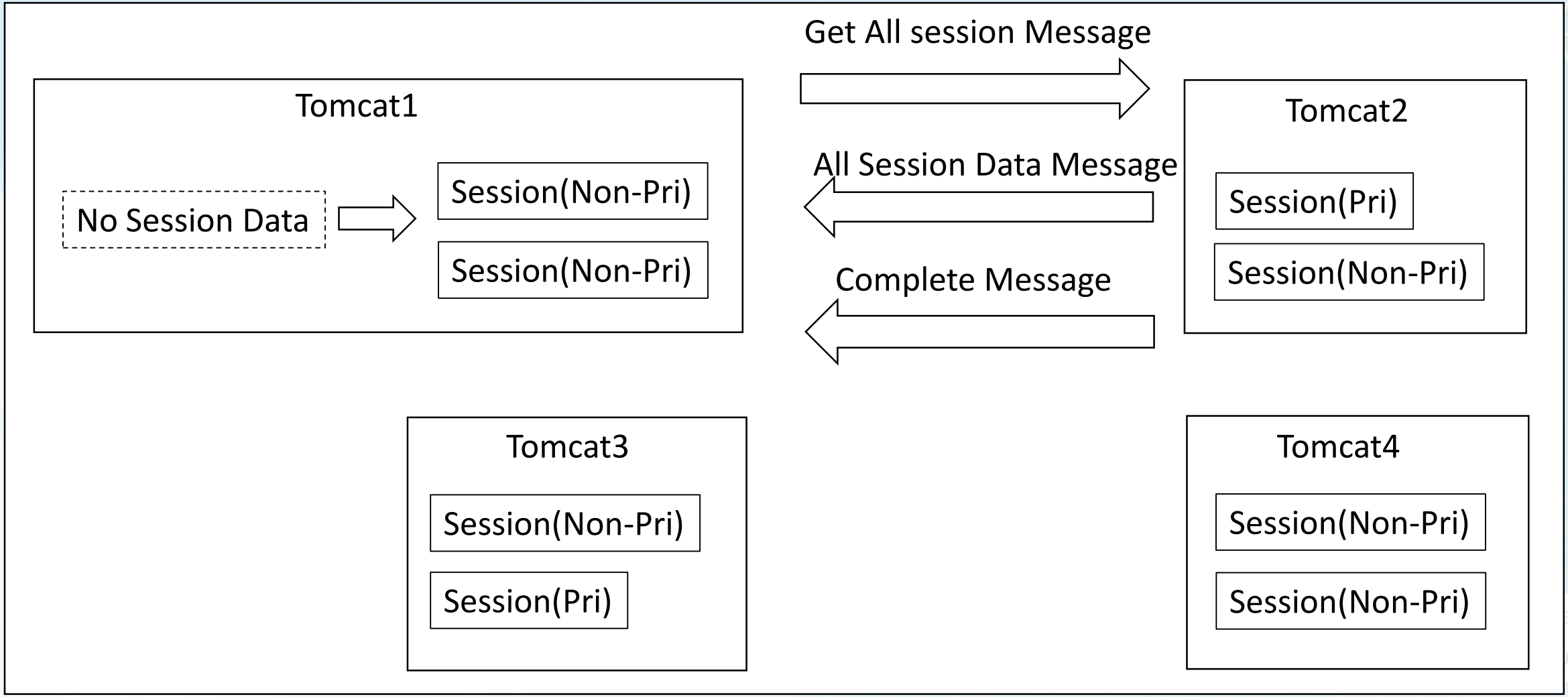
Behavior



Node Failure



Node Recovery

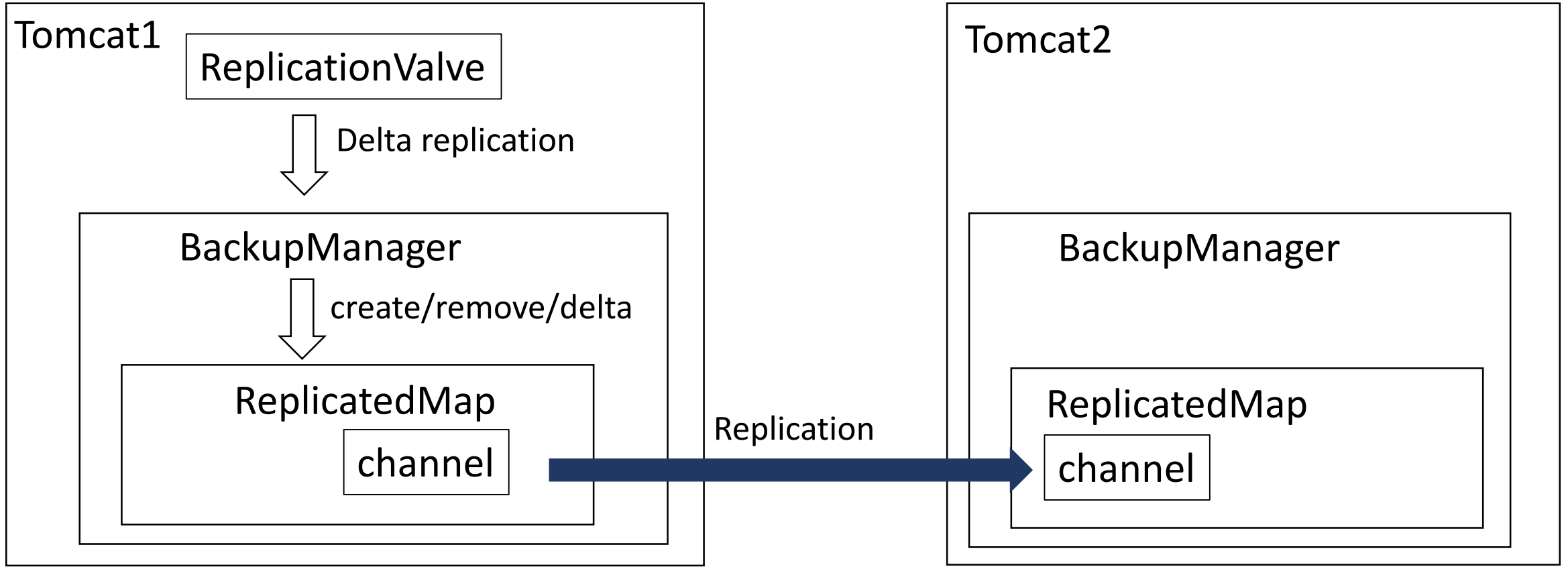


BackupManager

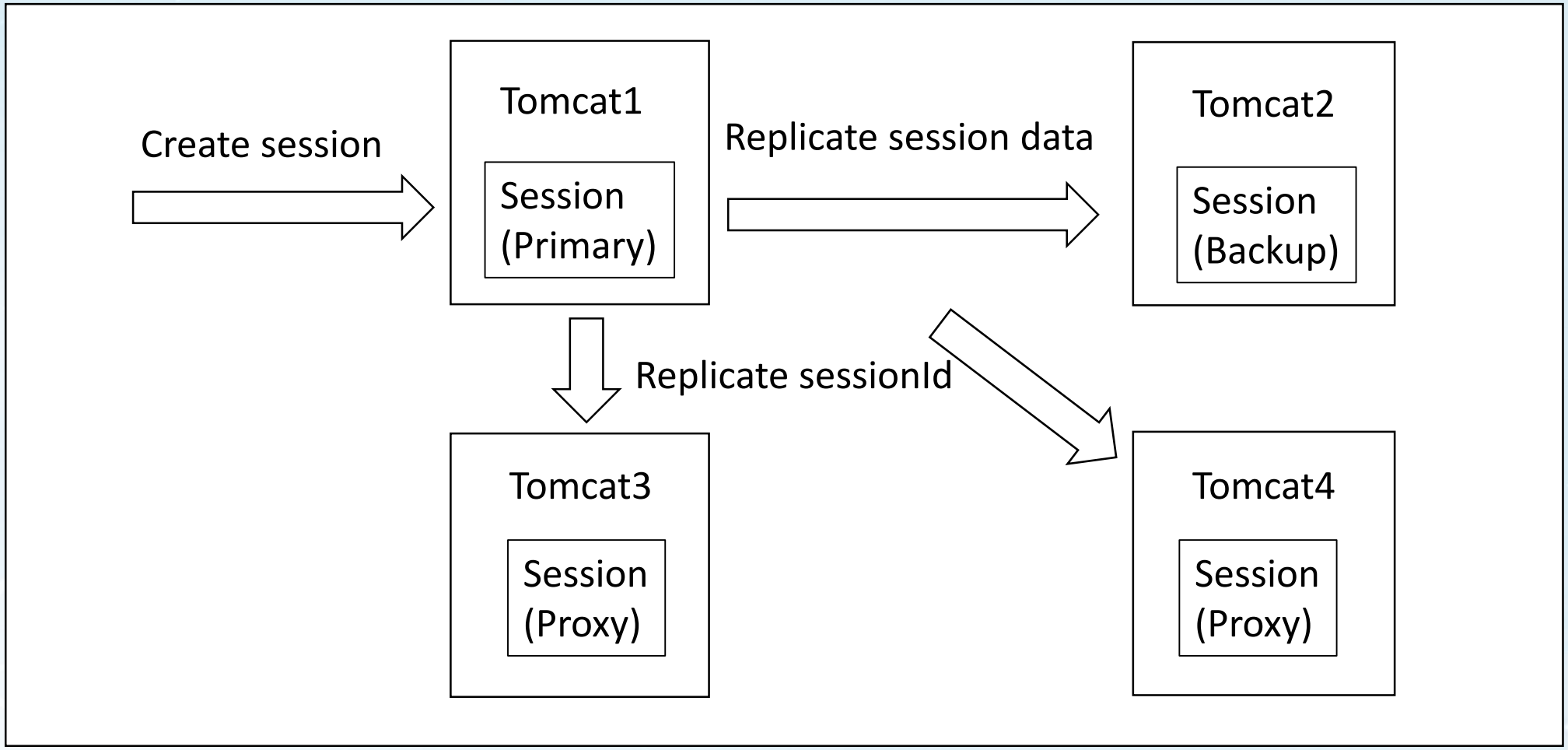
BackupManager

- Primary-Secondary session replication
- For large cluster

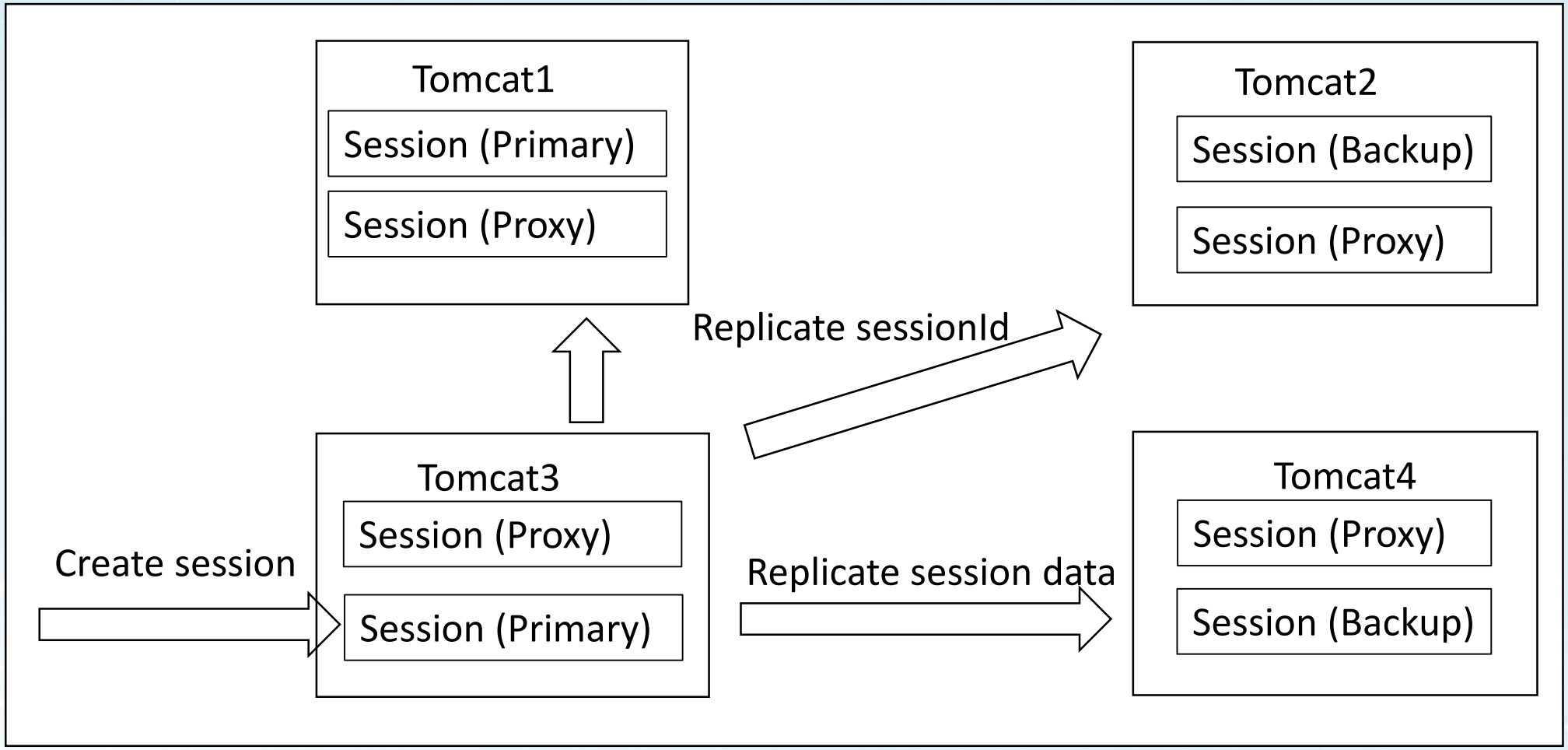
Architecture



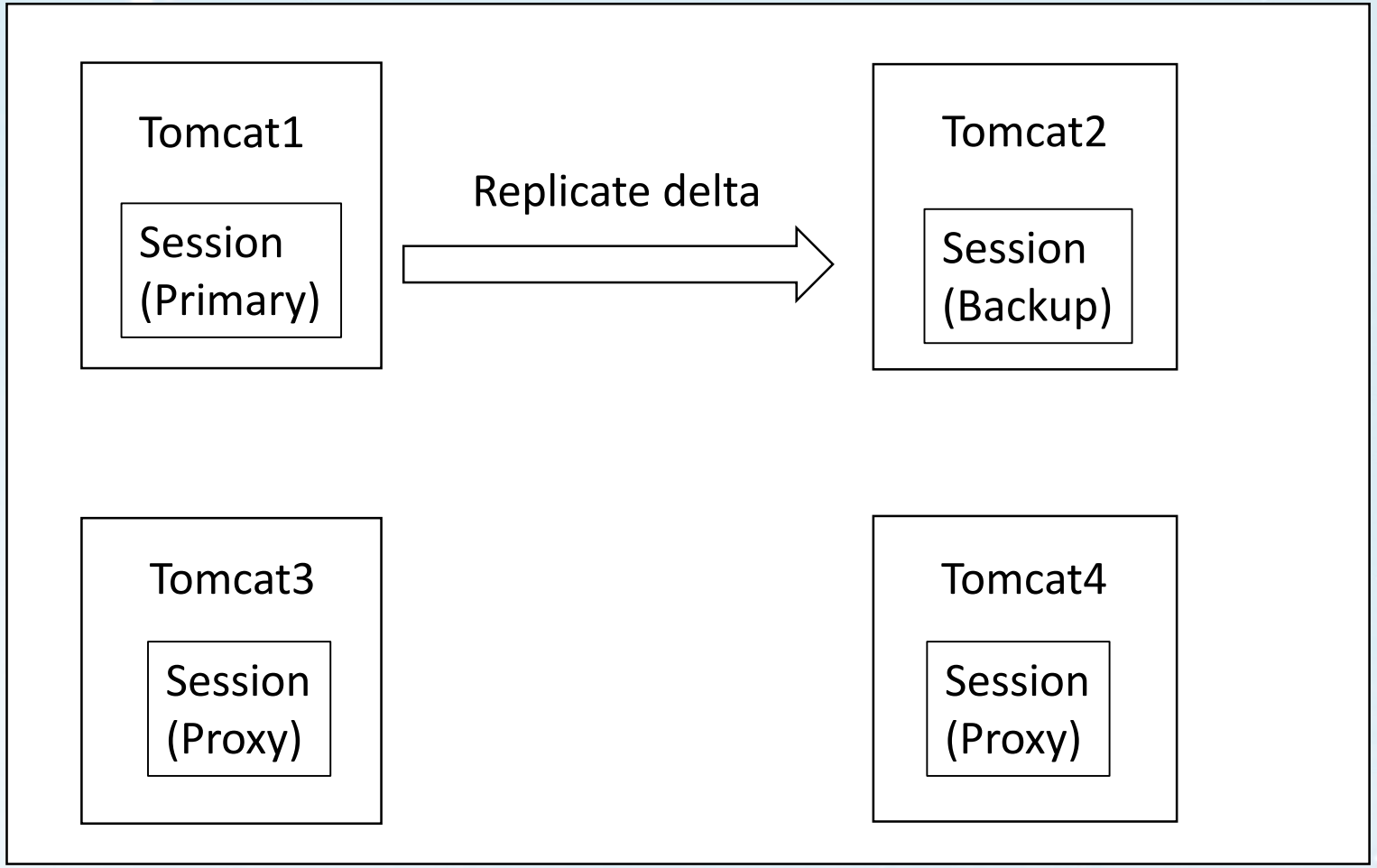
Behavior of Create



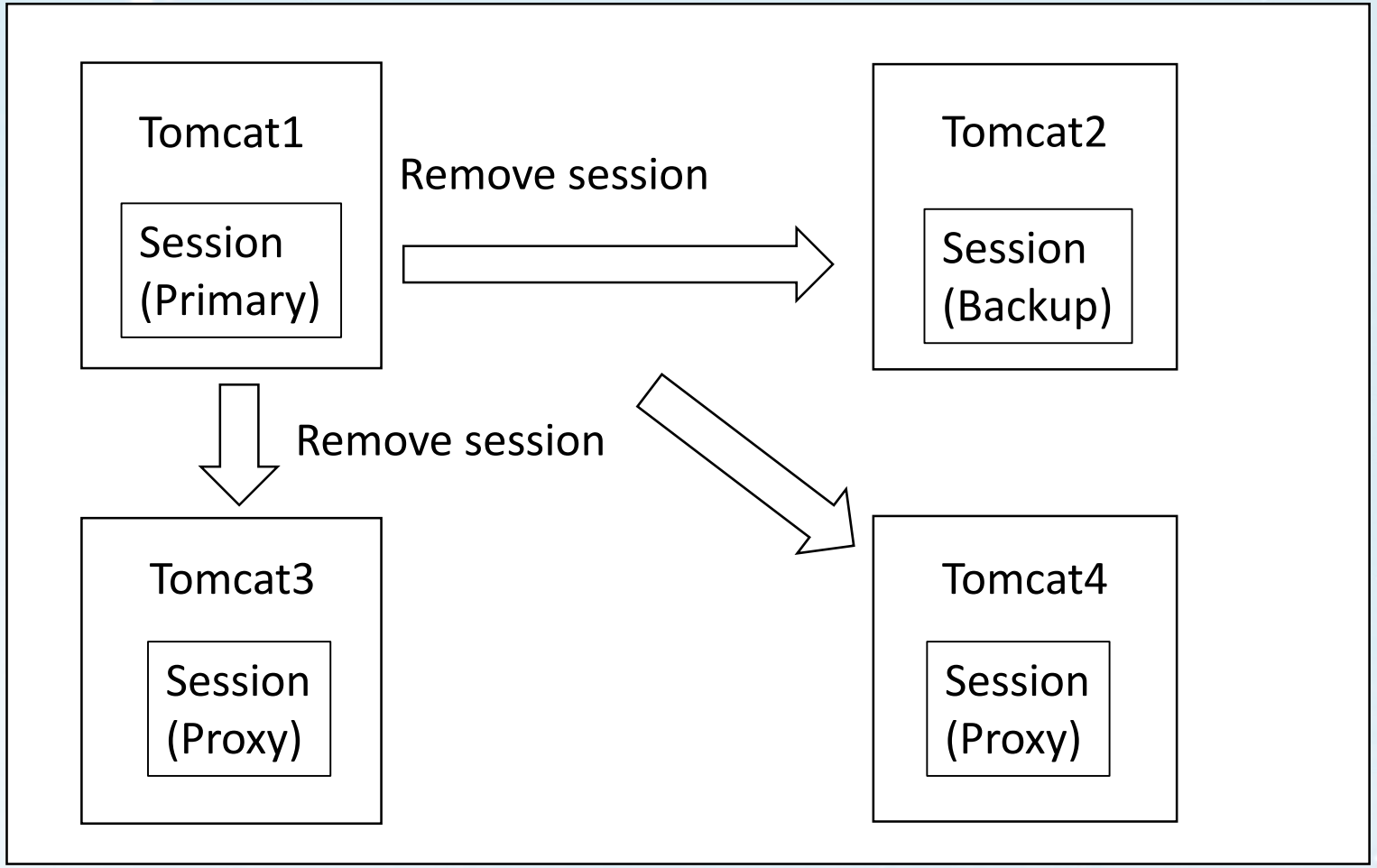
Behavior of Create



Behavior of Delta

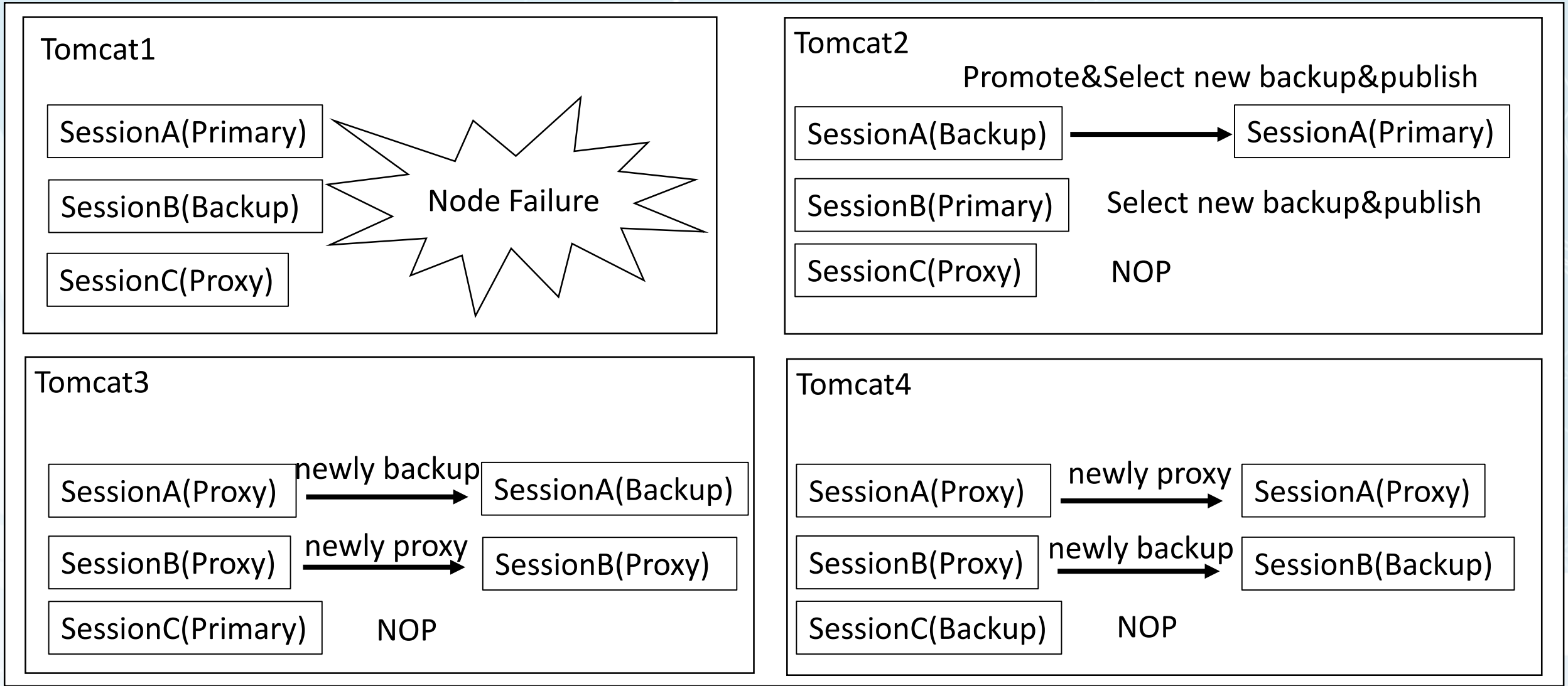


Behavior of Remove

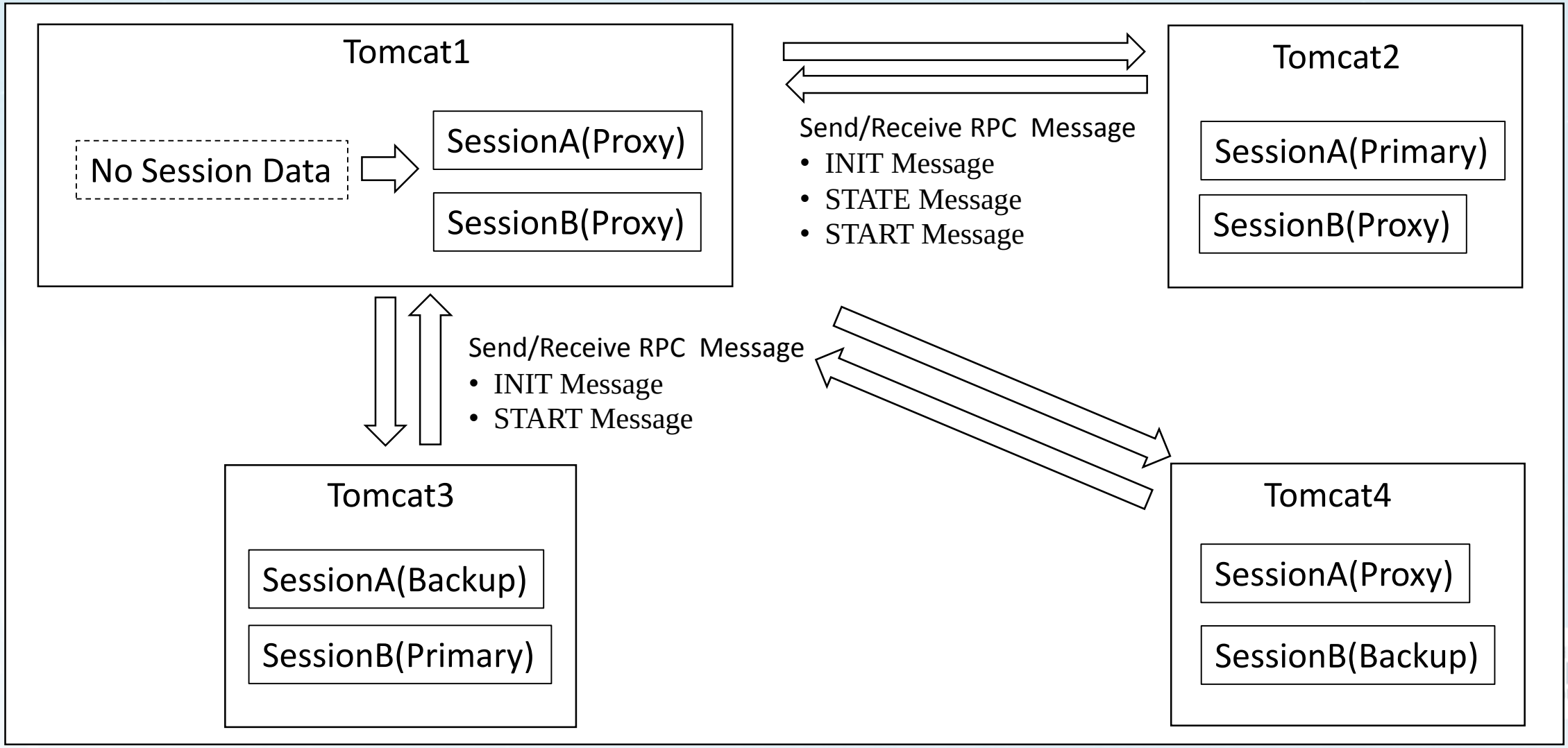


BackupManager

Node Failure



Node Recovery



Cluster Channel

Channel

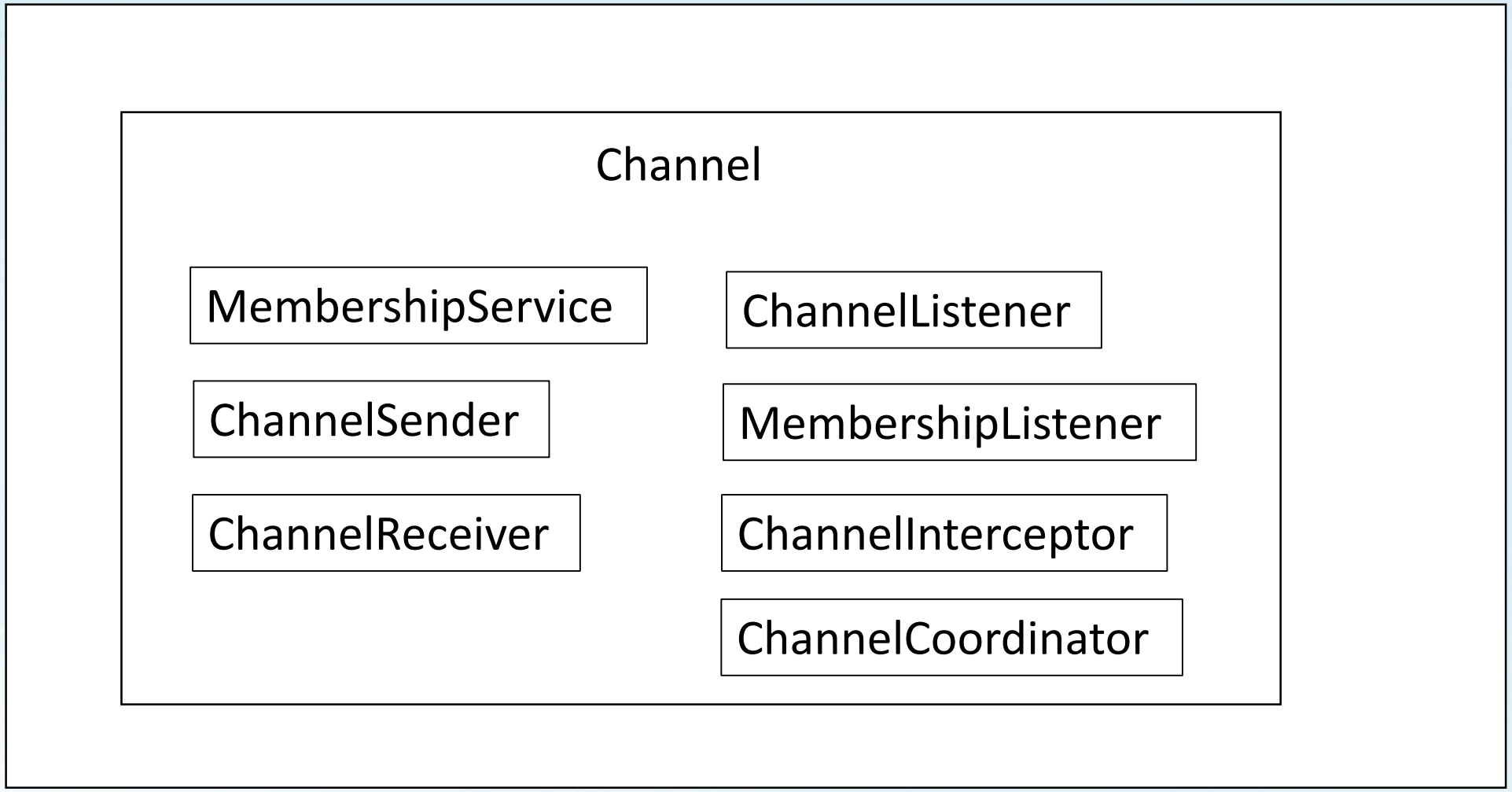
What is Channel ?

- Messaging & Grouping component

Responsibility

- Build Membership
- Send channel messages
- Receive channel messages

Channel Components



Channel Components

- Channel
  - The main component of channel
  - org.apache.catalina.tribes.group.GroupChannel only
- MembershipService
  - The component which build a cluster group
  - Start a multicast receivier thread and a multicast sender thread

Channel Components

- ChannelSender
  - Send channel messages to other nodes
  - Sender Queue size is specified in poolSize attribute
- ChannelReceiver
  - Receive channel messages from other nodes
  - Tuning of the maxThreads attribute depends on send option of channel message
    - Synchronization mode
      - Need to align maxThreads with poolSize
    - Asynchronous mode
      - do not need to align the maxThreads with poolSize.

Channel Components

- ChannelListener
  - Listen received channel messages
- MembershipListener
  - Listen add/remove of cluster members

Channel Components

- ChannelInterceptor
  - Intercept a channel message and a member detection
  - There are many configurable implementations
- ChannelCoordinator
  - Special ChannelInterceptor
  - Coordinates the ChannelInterceptors

Sample config

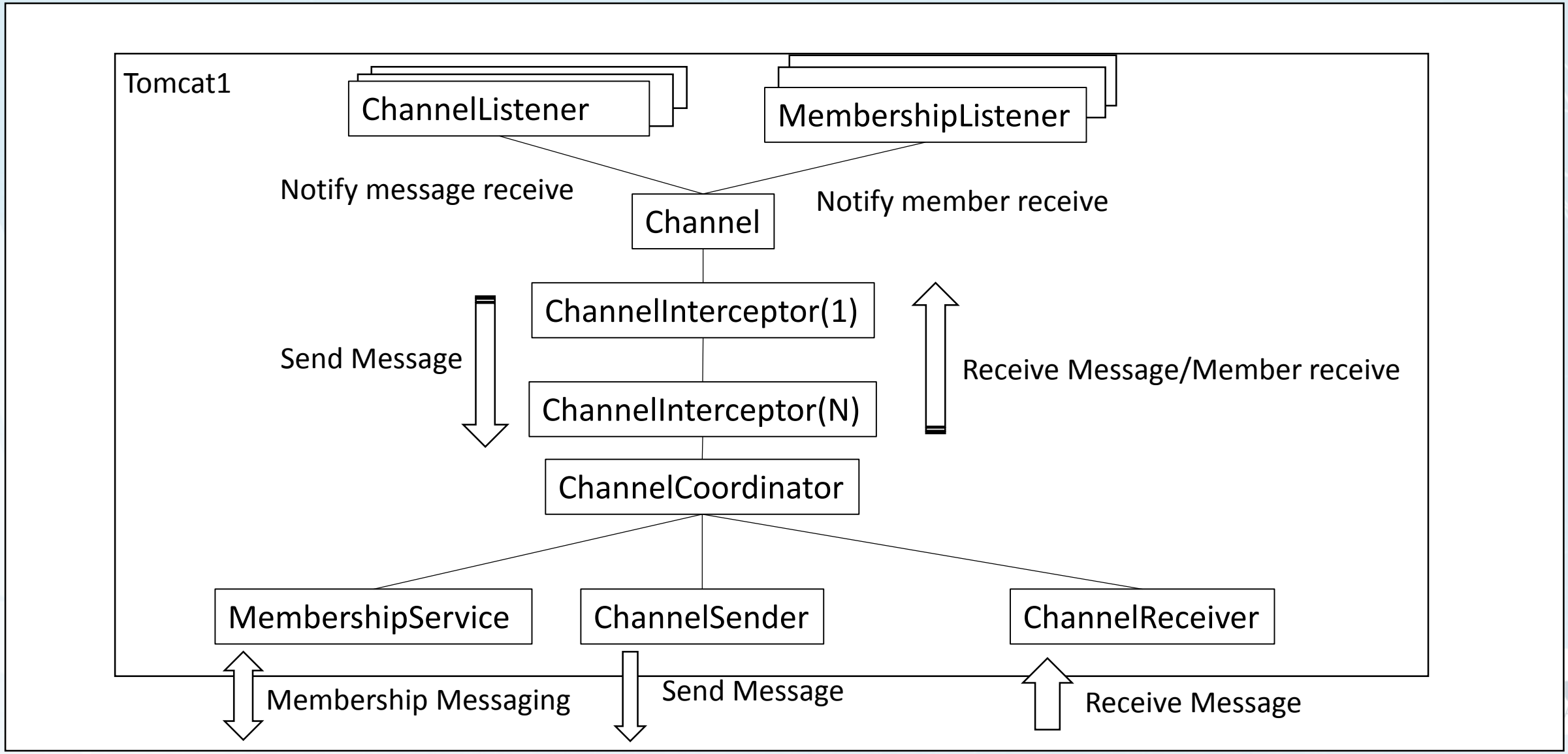

```
<Cluster className="org.apache.catalina.ha.tcp.SimpleTcpCluster">
 ... ...
  <Channel className="org.apache.catalina.tribes.group.GroupChannel">
    <Membership className="org.apache.catalina.tribes.membership.McastService"
       address="229.0.0.61“ port="45564“ frequency="500“ dropTime="4000" />

    <Receiver className="org.apache.catalina.tribes.transport.nio.NioReceiver"
       address=“auto“ port=”4004“ autoBind=”100” maxThreads=“25”/>

    <Sender className="org.apache.catalina.tribes.transport.ReplicationTransmitter">
      <Transport className="org.apache.catalina.tribes.transport.nio.PooledParallelSender"
         timeout="5000“  poolSize="25"/>
    </Sender>

    <Interceptor className="org.apache.catalina.tribes.group.interceptors.MessageDispatch15Interceptor" />
    <Interceptor className="org.apache.catalina.tribes.group.interceptors.TcpFailureDetector"/>
  </Channel>
 ... ...
</Cluster>
```

Channel Architecture



ChannelInterceptor

TCPFailureDetector

- Main Features
  - Intercept memberDisappeared events.
  - Member check when send errors.

- Other Features
  - Manage membership in static membership

ChannelInterceptor

MessageDispatchInterceptor

- Asynchronous send message
- Use Thread Pooling
  - maxThreads
  - maxSpareThreads
- You must set the send options asynchronous.
  - Cluster’s channelSendOptions to asyn mode(8)
  - BackupManager’s mapSendOptions to asyn mode(8)

ChannelInterceptor

StaticMembershipInterceptor

- The static membership instead of multicast

- Make sure
  - Disable multicast membership
    - Cluster’s channelStartOptions attribute = 3
  - Enable TcpPingInterceptor for nodes failure detection
  - Enable TcpFailureDetector for membership management
  - The order is

    TcpPingInterceptor
    TcpFailureDetector
    StaticMembershipInterceptor

ChannelInterceptor


```
<Interceptor className="org.apache.catalina.tribes.group.interceptors.TcpPingInterceptor"/>
<Interceptor className="org.apache.catalina.tribes.group.interceptors.TcpFailureDetector"/>
<Interceptor className=
      "org.apache.catalina.tribes.group.interceptors.StaticMembershipInterceptor">
 <LocalMember className="org.apache.catalina.tribes.membership.StaticMember"
    uniqueId="{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,1}"/>
  <Member className="org.apache.catalina.tribes.membership.StaticMember"
     port="4010" host="hostA"
     uniqueId="{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,2}" />
  <Member className="org.apache.catalina.tribes.membership.StaticMember"
     port="4010" host="hostB"
     uniqueId="{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,3}" />
</Interceptor>
```

ChannelInterceptor

ThroughputInterceptor

- Measuring the send and receive of channel messages

DomainFilterInterceptor

- Filter the members that join cluster group by the domain

Monitoring Cluster

Monitoring Cluster

How to monitor Tomcat cluster.

- Logging
- JMX

Monitoring Cluster

Monitoring your Cluster with Logging

- To track channel messages for Debugging
  - Use the Tomcat JULI
  - Key: org.apache.catalina.tribes.MESSAGES
  - Level: FINEST

```
# FOR DEBUG
10catalina.org.apache.juli.AsyncFileHandler.formatter = org.apache.juli.VerbatimFormatter
10catalina.org.apache.juli.AsyncFileHandler.level = FINEST
10catalina.org.apache.juli.AsyncFileHandler.directory = ${catalina.base}/logs
10catalina.org.apache.juli.AsyncFileHandler.prefix = MESSAGES.
10catalina.org.apache.juli.AsyncFileHandler.bufferSize = -1
org.apache.catalina.tribes.MESSAGES.level = FINEST
org.apache.catalina.tribes.MESSAGES.handlers = 10catalina.org.apache.juli.AsyncFileHandler
```

Monitoring your Cluster with Logging

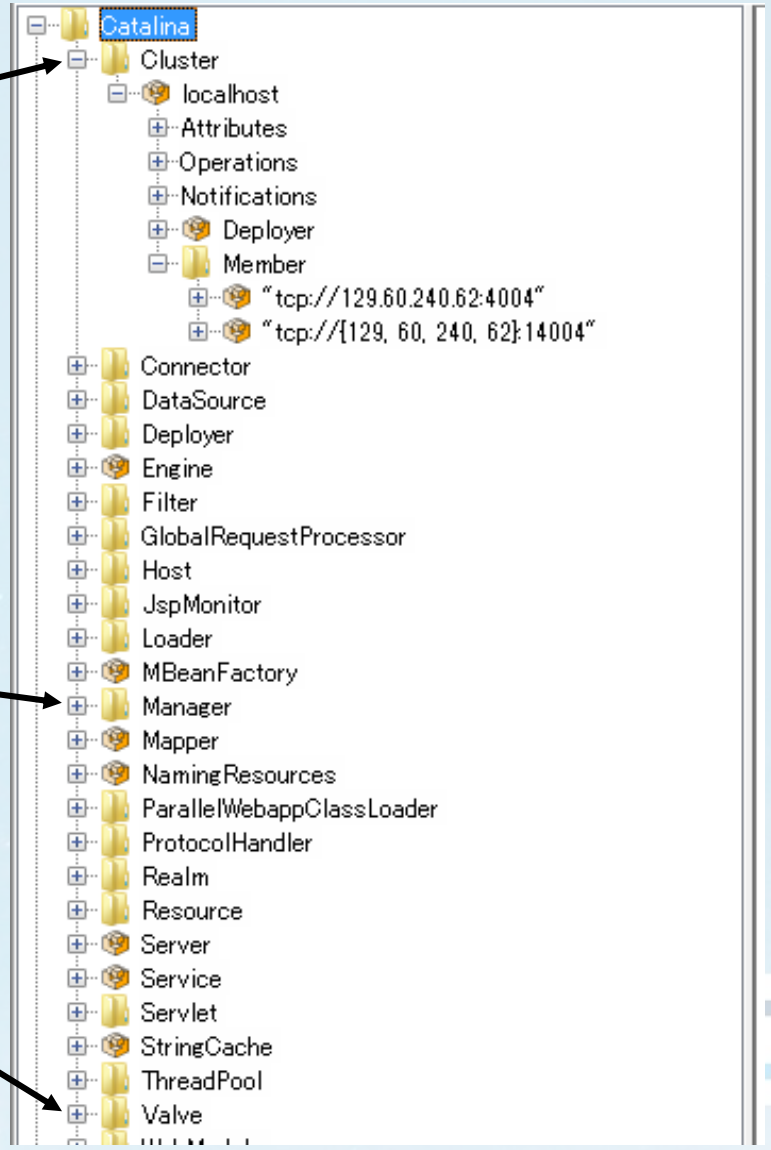
- ThroughputInterceptor
  - Report the throughput statistics.
  - Default interval is to report every 10000 messages.
  - Key: org.apache.catalina.tribes.group.interceptors.ThroughputInterceptor
  - Level: INFO

```
org.apache.catalina.tribes.group.interceptors.ThroughputInterceptor.report ThroughputInterceptor Report[
        Tx Msg:60,024 messages
        Sent:53.30 MB (total)
        Sent:53.31 MB (application)
        Time:16.84 seconds
        Tx Speed:3.16 MB/sec (total)
        TxSpeed:3.16 MB/sec (application)
        Error Msg:0
        Rx Msg:60,048 messages
        Rx Speed:0.23 MB/sec (since 1st msg)
        Received:53.28 MB]
```

Monitoring Cluster

Monitoring your Cluster with JMX

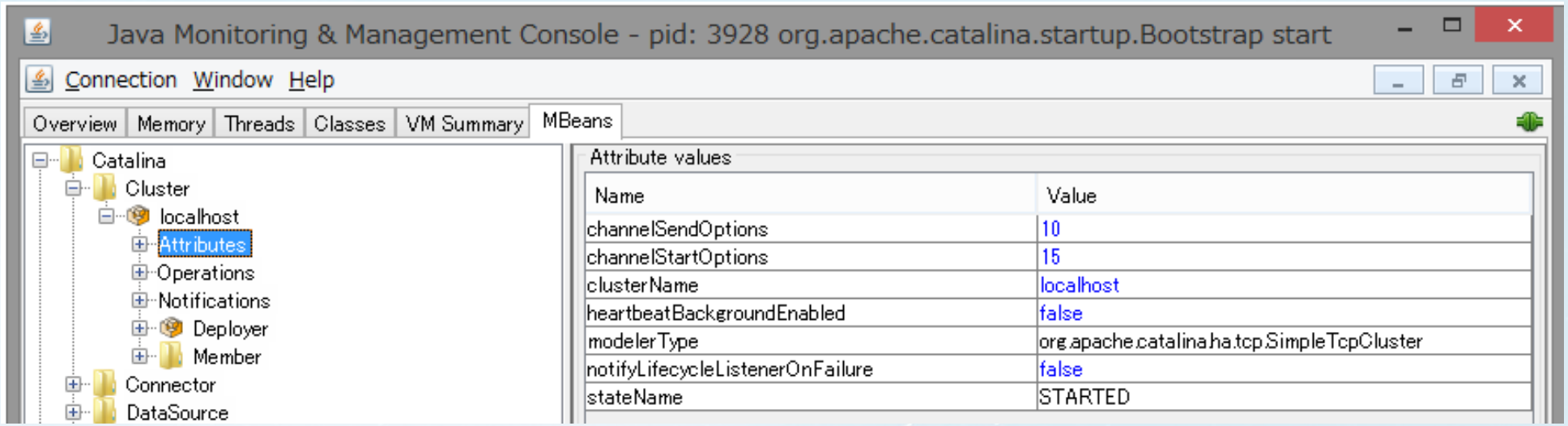
- Catalina Domain
  - Cluster Mbeans
    - Cluster Mbean
    - Deployer Mbean
    - Member MBeans
  - (Cluster)Manager Mbean
  - (Cluster)Valve MBean



Monitoring Cluster

Cluster Mbean

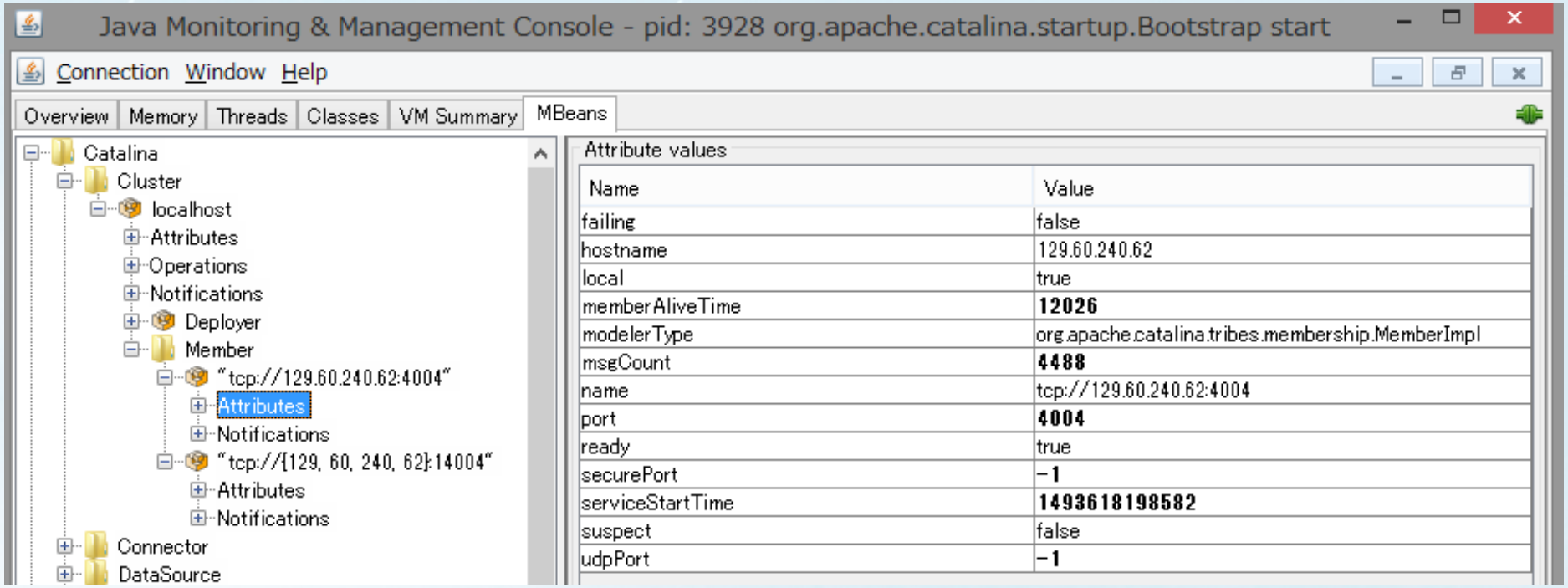
- Cluster settings



Monitoring Cluster

Member Mbeans

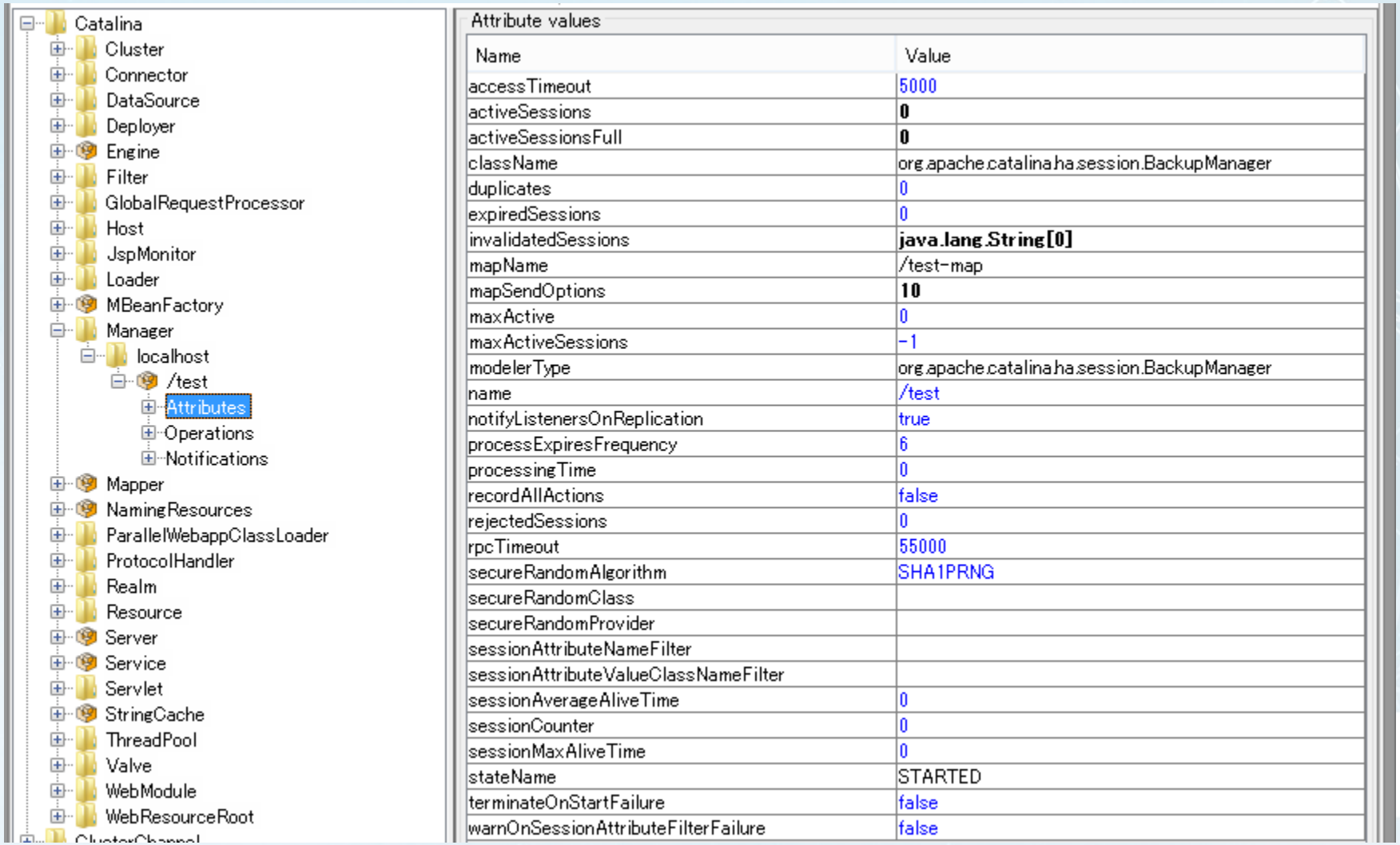
- All Cluster members that have been joining same cluster group



Monitoring Cluster

(Cluster)Manager Mbean

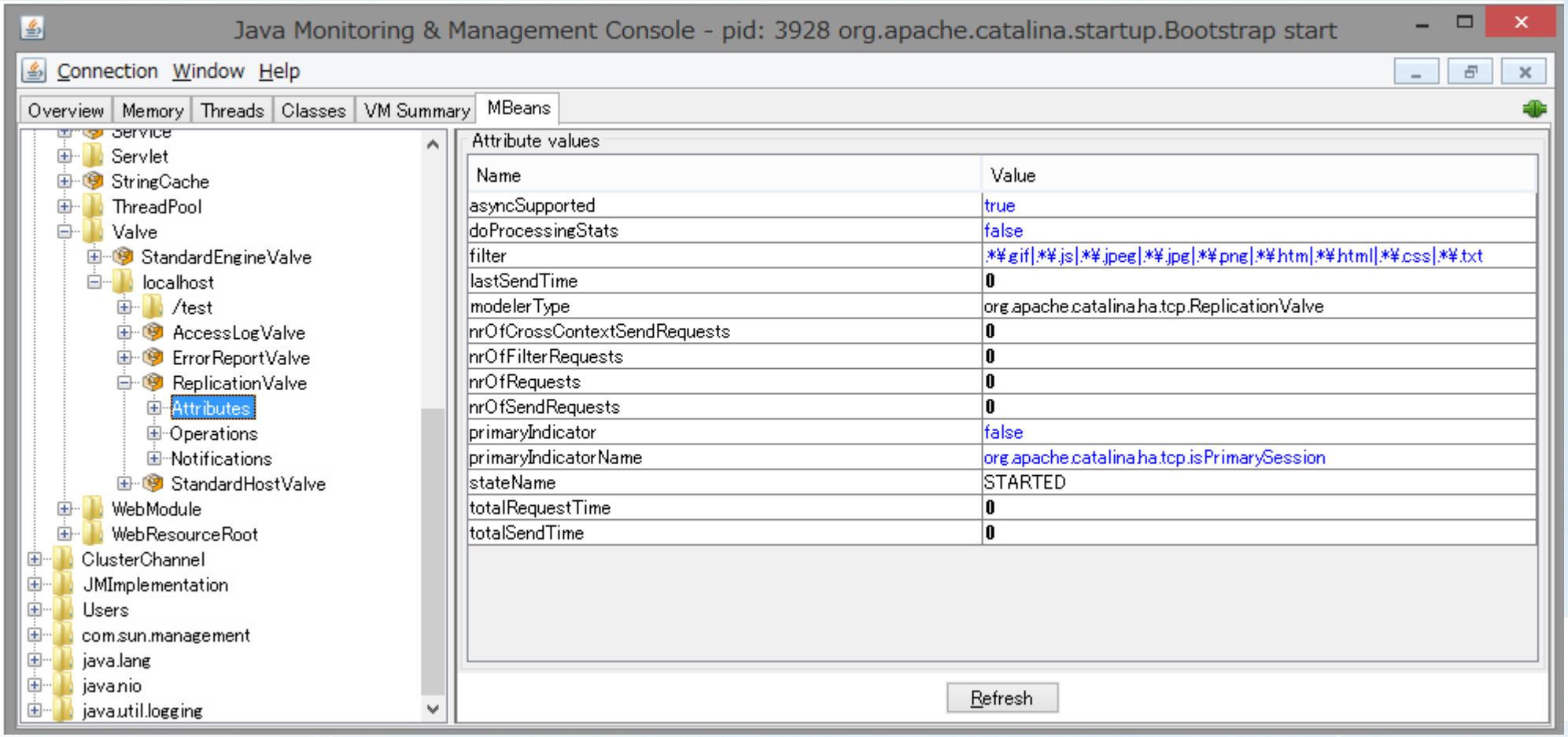
- Session Information



Monitoring Cluster

(Cluster)Valve MBean

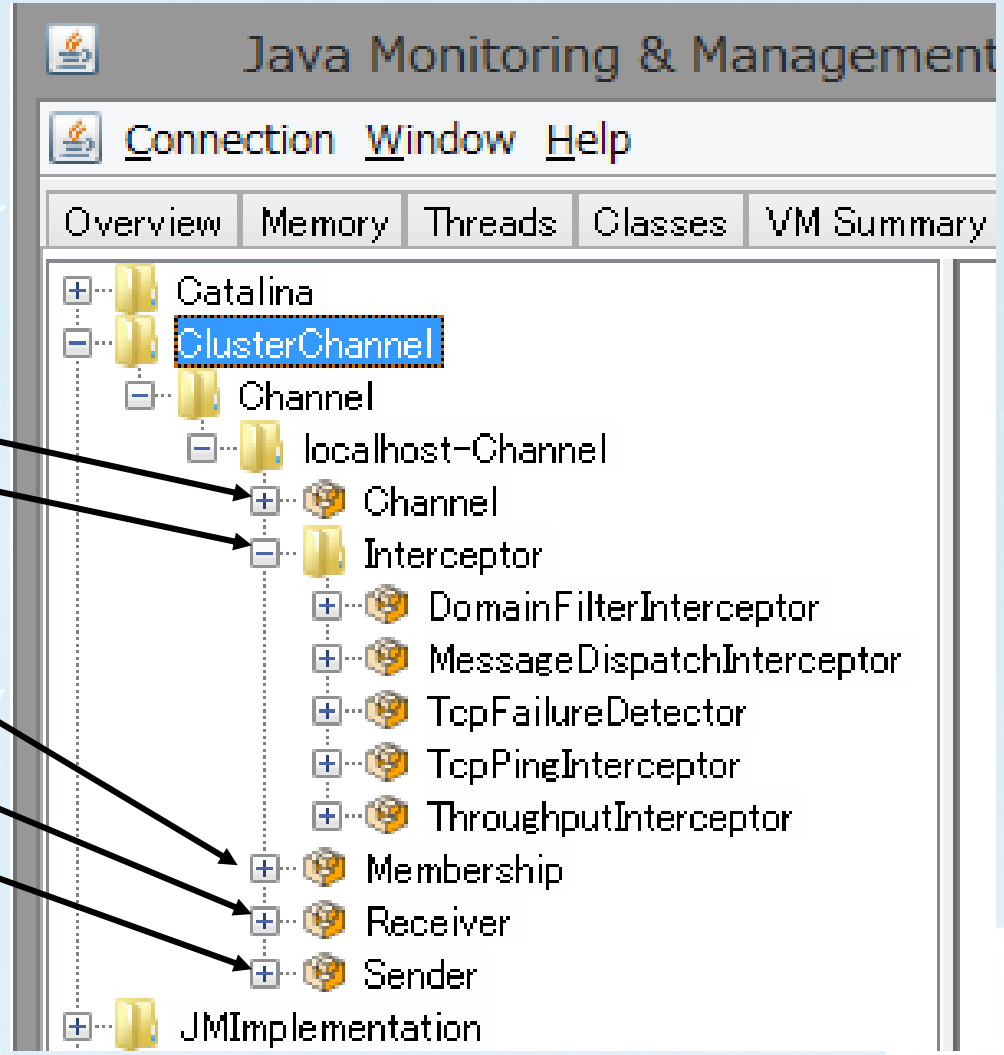
- ReplicationValve settings



Monitoring Cluster

Monitoring your Cluster with JMX

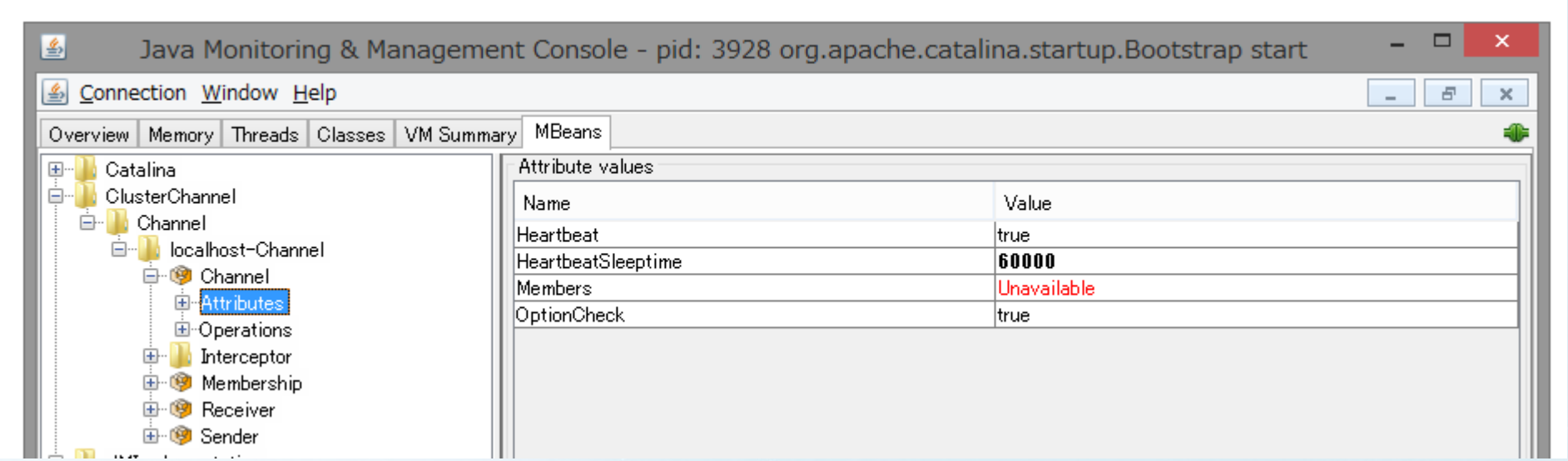
- ClusterChannel Domain
  - Channel Mbeans
    - Channel Mbean
    - Interceptor Mbeans
    - Membership Mbean
    - Receiver Mbean
    - Sender MBean



Monitoring Cluster

Channel Mbean

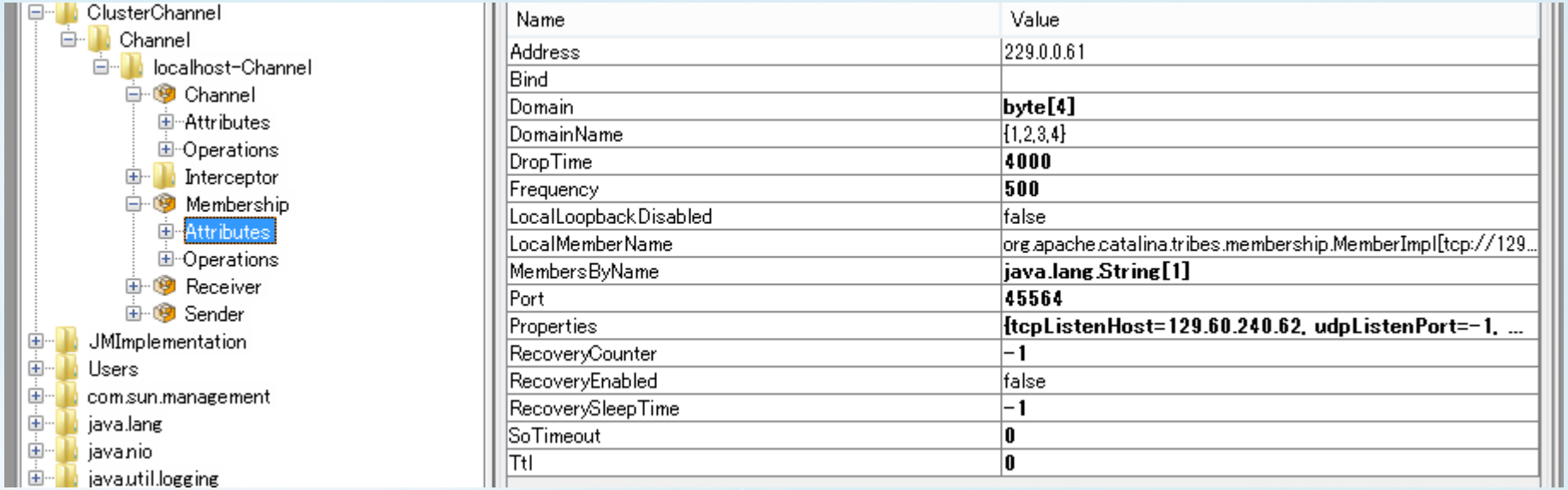
- Channel settings



Monitoring Cluster

Membership Mbean

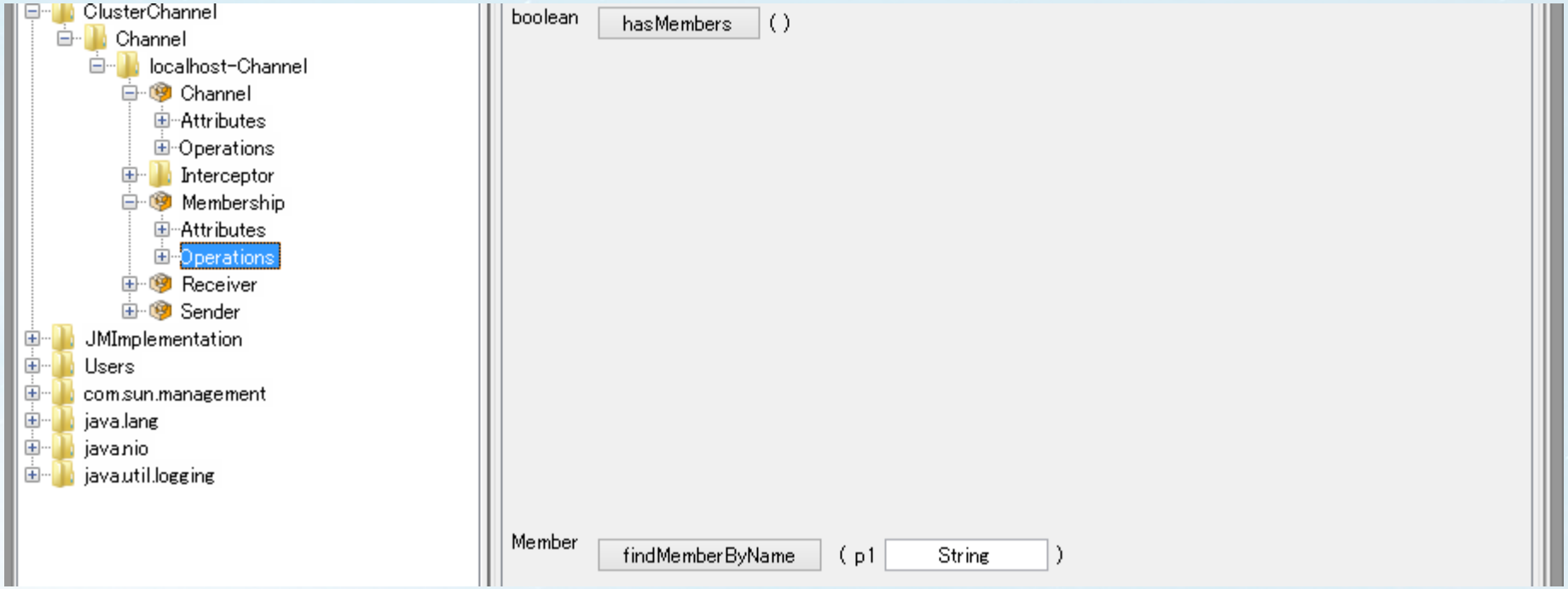
- Membership Settings



Monitoring Cluster

Membership Mbean

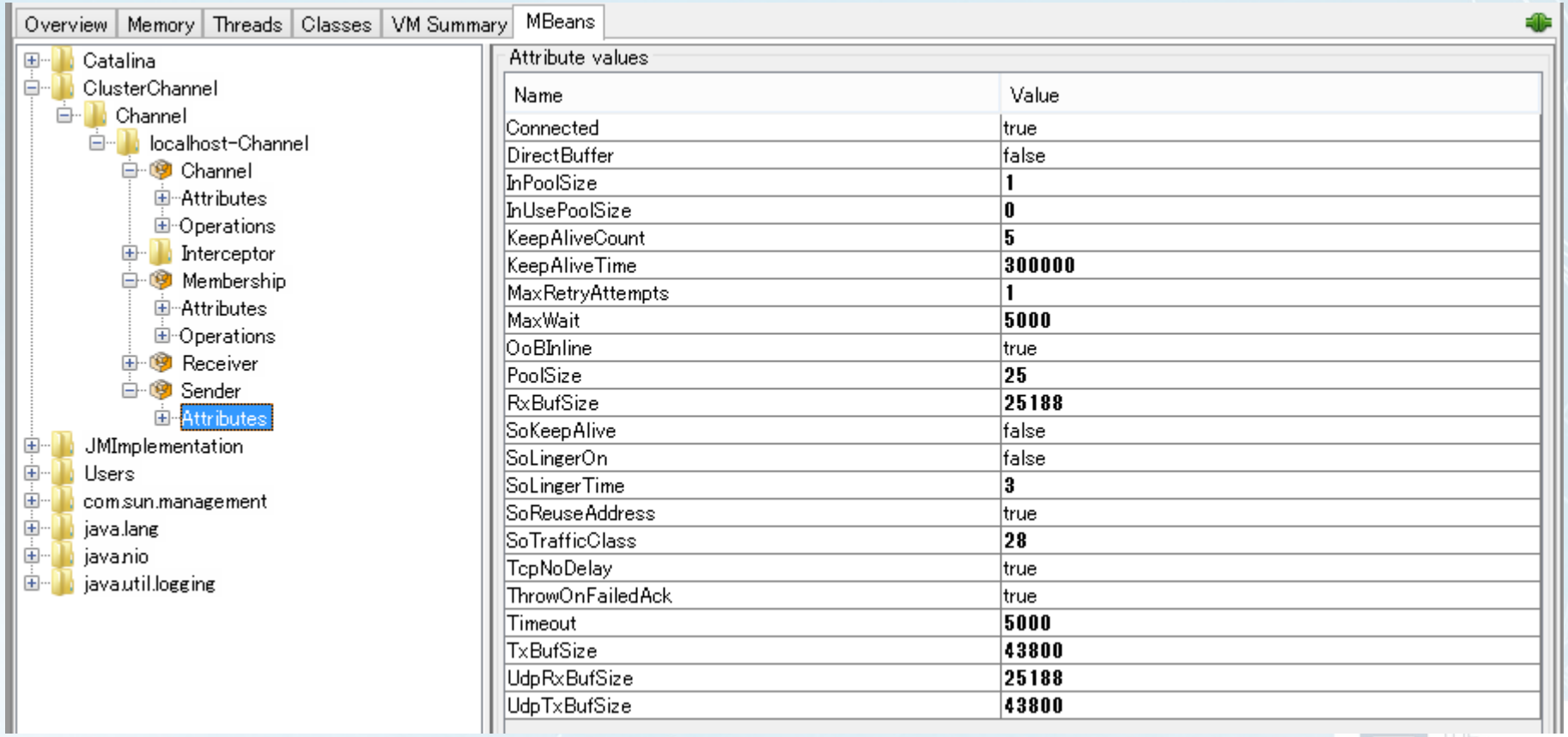
- Membership Operations



Monitoring Cluster

Sender MBean

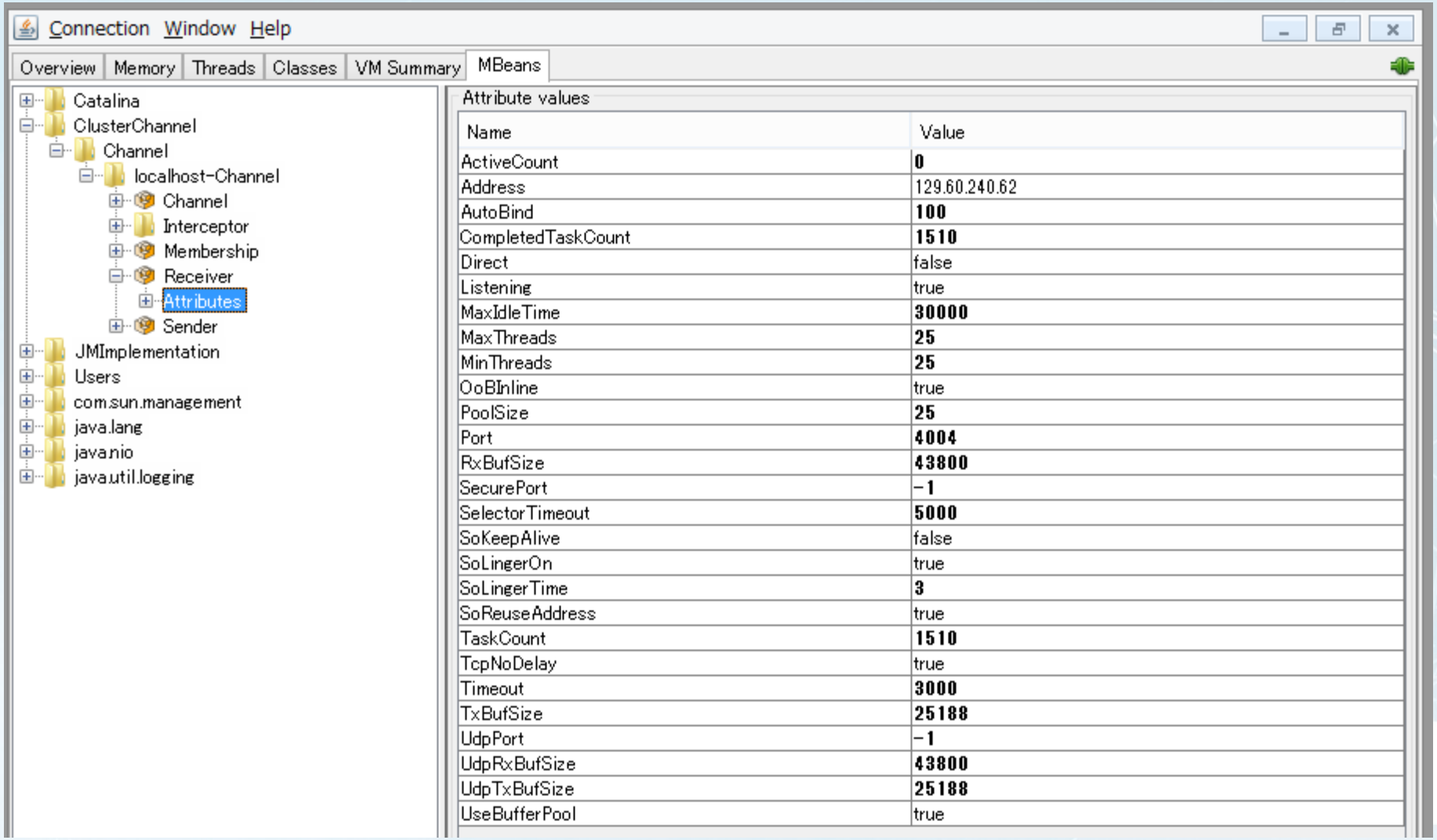
- Settings and Stats info



Monitoring Cluster

Receiver Mbean

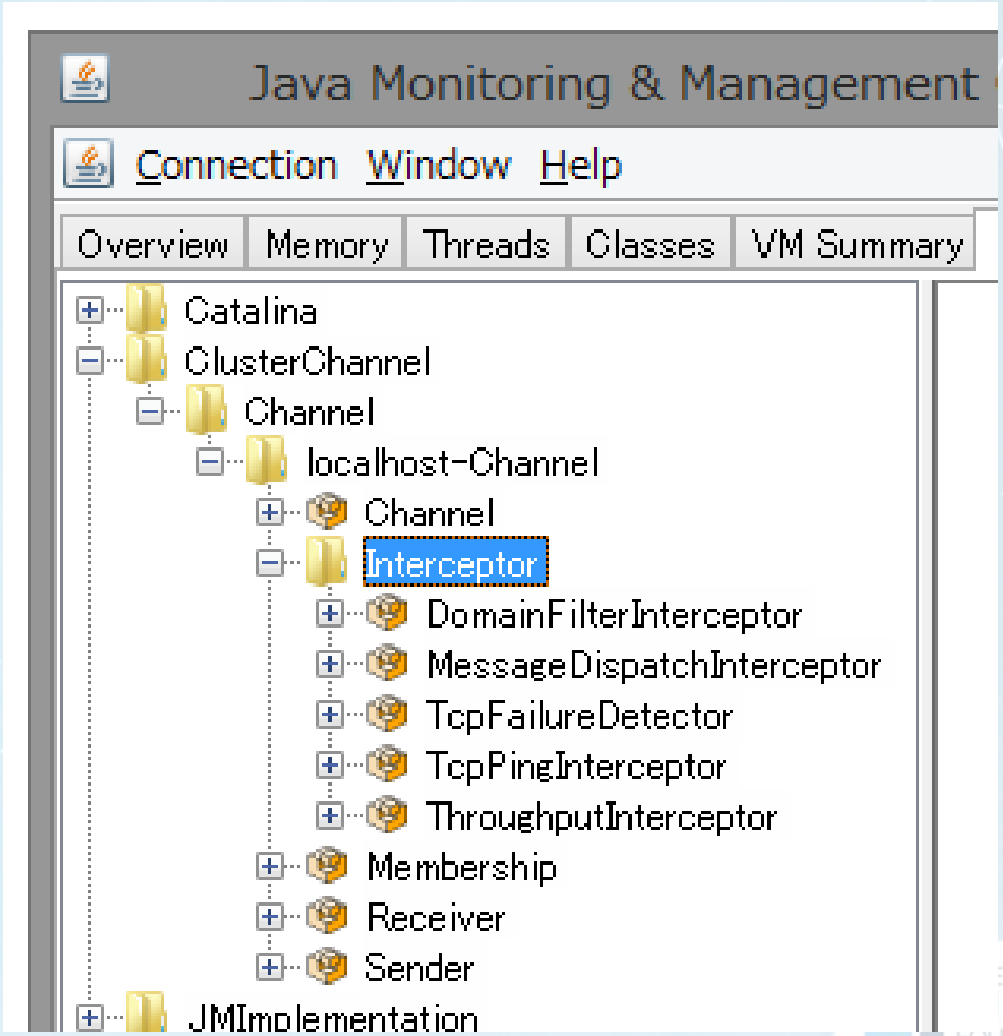
- Settings and Stats



Monitoring Cluster

Interceptor Mbeans

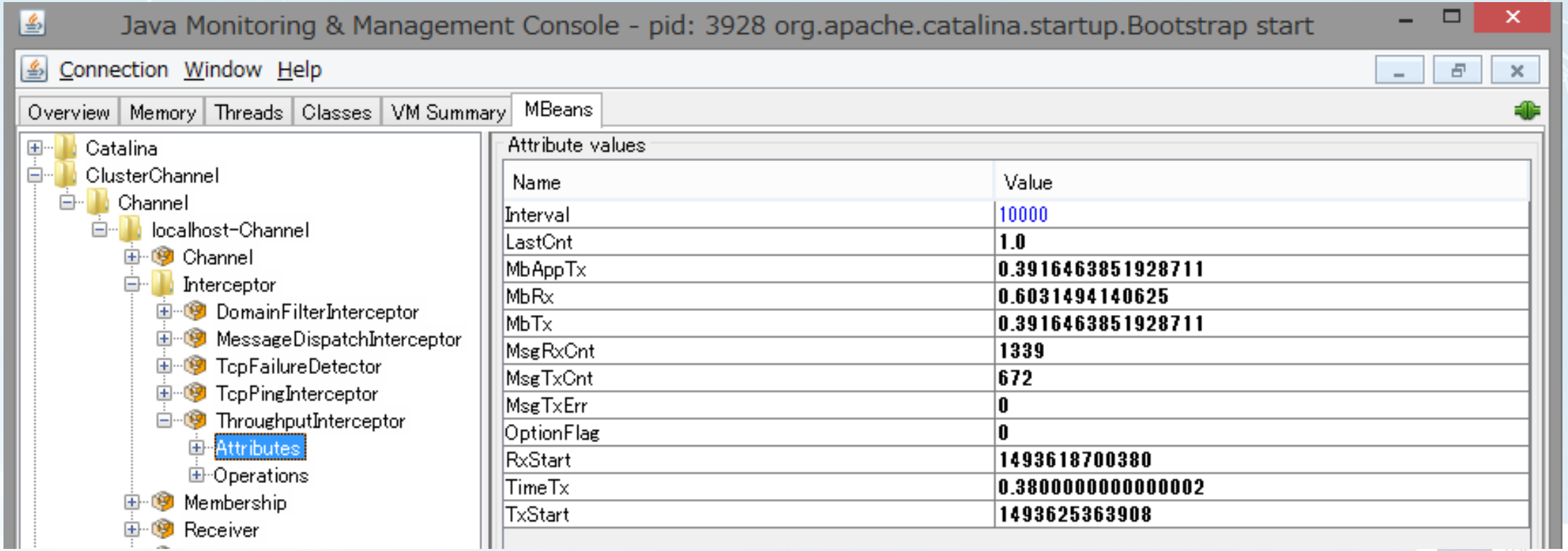
- All interceptor that are used in channel
- Implements MBeans of all commonly used Interceptors but It does not implement all Interceptor MBeans.



Monitoring Cluster

ThroughputInterceptor

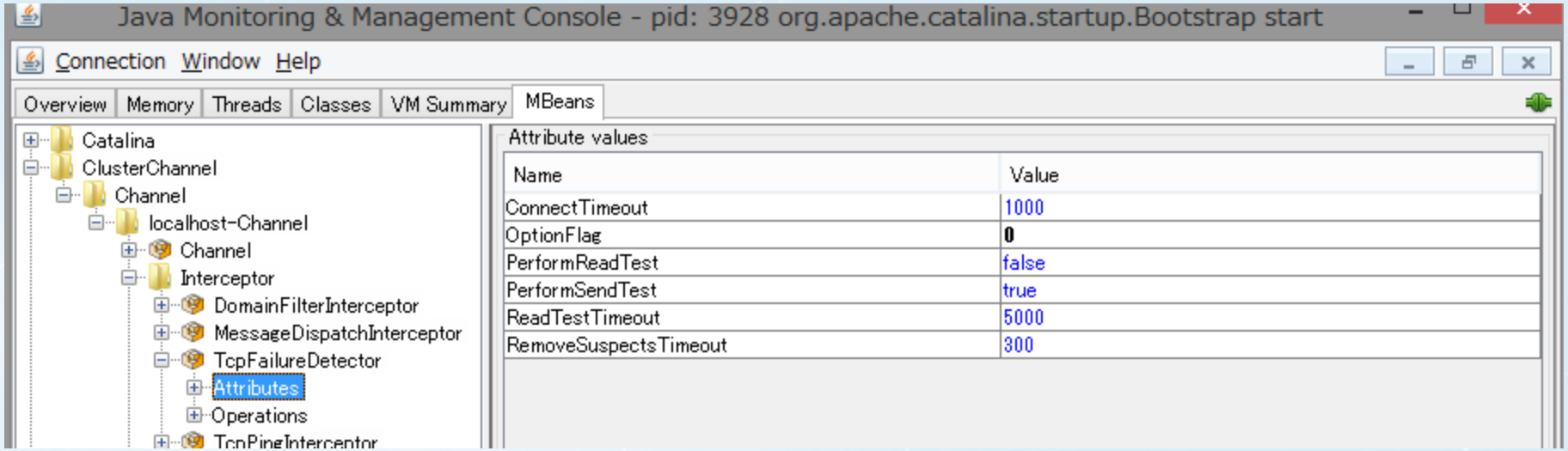
- The throughput statistics

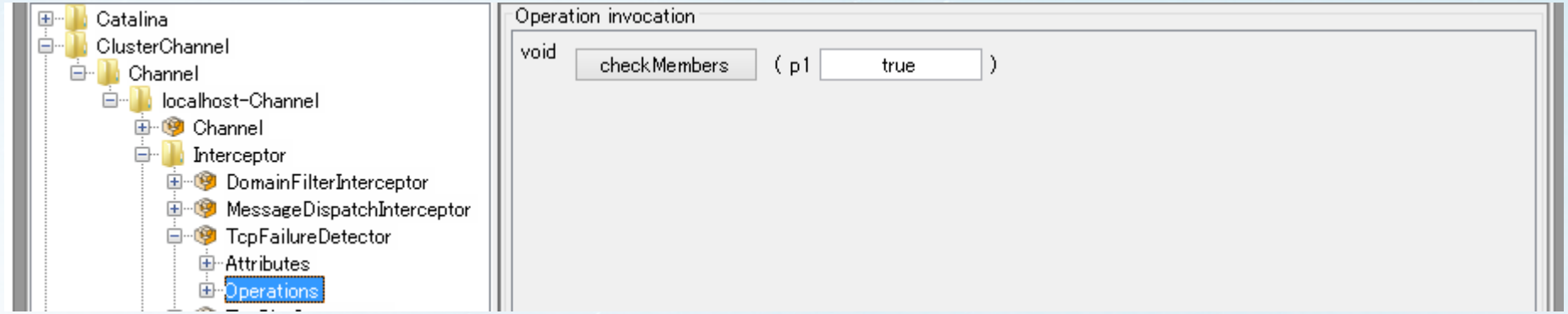


Monitoring Cluster

TcpFailureDetector

- Settings and Member check by TCP

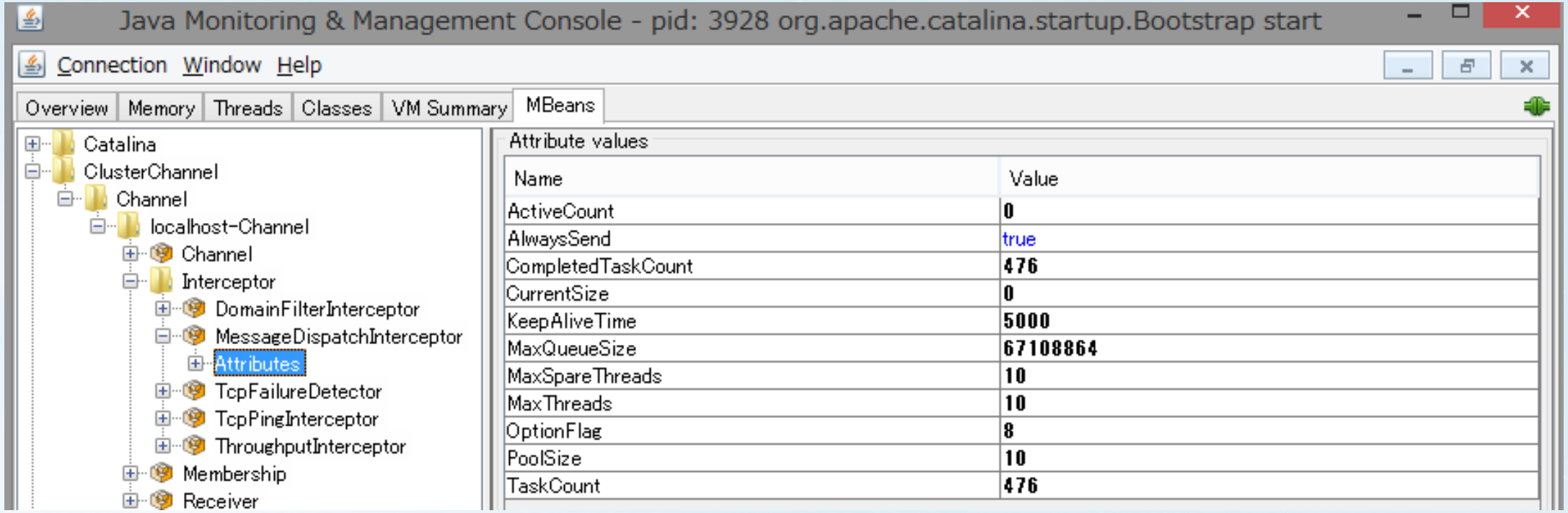




Monitoring Cluster

MessageDispatchInterceptor

- Settings and Stats info



Monitoring Cluster

It's important to note that

- Channel Mbeans are supported in 9.0.0.M20 and later
- This feature has not back-ported into Tomcat8.5 yet.
- The ReplicationMap MBean is T.B.D

Questions?

Thank You

APACHECON North America
MAY 16-19, 2017
MIAMI, FL

Tomcat Cluster

Keiichi Fujino